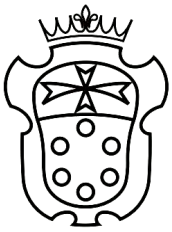


SCUOLA
NORMALE
SUPERIORE
PISA

CLASSE DI SCIENZE
Corso di Perfezionamento in Fisica

Tesi di perfezionamento

Search for the Standard Model Higgs boson decaying to b quarks with the CMS experiment

Candidato:
Silvio Donato

Relatori:
Prof. Luigi Rolandi
Prof. Andrea Rizzi

Anno Accademico 2015-2016

Introduction

The first run of the LHC has been a great leap forward in our understanding of Standard Model (SM) high energy physics [1–3] with the discovery of the Higgs boson [4, 5] performed by the CMS [6] and ATLAS [7] collaborations and the subsequent measurements of its properties. All properties measured so far are compatible with the SM predictions. Still, one of the main missing pieces in the Higgs boson puzzle is the direct observation of the Higgs boson decaying to b quarks. Although the branching ratio of the decay $H \to bb$ predicted by the SM is about 58%, the observation of this channel is still missing because of the poor di-jet mass resolution compared to di-photons ($H \to \gamma\gamma$) or 4-leptons ($H \to 4\ell$) final states, and because the sensitivity to the $gg \to H \to bb$ channel is suppressed by the large QCD multijet background.

In April 2015, after the Long Shutdown 1, LHC delivered the first proton-proton collisions at the unprecedented energy of $\sqrt{s} = 13\,\text{TeV}$. In the next few years, LHC is expected to increase its instantaneous luminosity up to $2 \cdot 10^{34}\,\text{cm}^{-2}\,\text{s}^{-1}$, almost three times the peak luminosity of Run 1. With this luminosity, we will observe the decay $H \to bb$ during Run 2 if performing triggers at high luminosity allow to collect signal events with the largest acceptance and the highest efficiency.

This thesis describes the improvements on the High Level Trigger that I developed during the Long Shutdown 1 for the search of the $H \to bb$ decay with the CMS experiment. A key point of these triggers is the possibility to tag online the presence of a high momentum b quark in the final state. The online b tagging has been improved and made faster using a dedicated version of tracking that reconstructs only the tracks that are compatible with the primary vertex and close to the jet direction. This version of online tracking matching the primary vertex can run up to a rate of 100 kHz and is used also to detect and reject pile-up jets, improving the resolution on the missing transverse energy evaluated on calorimetric jets. I exploited these new tools to design triggers dedicated to the search for the channel $ZH \to \nu\nu bb$, $H \to bb$ produced through the vector boson fusion process, and a resonance decaying to $HH \to 4b$, enlarging significantly the trigger acceptance of these searches.

The last part of this thesis presents the search for the $ZH \to \nu\nu bb$ with the CMS experiment performed using the 13 TeV data collected in 2015 by the triggers that

I developed. The analysis strategy takes advantage of some improvements compared to the Run 1 analysis, like a new signal extraction fit and a new treatment of the multijet background. The analysis has been performed using $2.32\,\text{fb}^{-1}$ of data collected in 2015 and the expected results after the whole 2016 data taking ($\sim 20\,\text{fb}^{-1}$) have been estimated.

Outline of this document

Chapter 1 describes the theory of the Higgs boson in the framework of the SM and gives an overview of the physics of the Higgs boson and of its search performed by the CMS and ATLAS collaborations during the LHC Run 1.

Chapter 2 introduces the CMS experiment at the LHC and the event reconstruction used by CMS. My contribution to the event reconstruction is an improvement of the tracking in the core of energetic jets, described in Appendix A. The last section of this Chapter will present the trigger system of CMS, with special attention to the High Level Trigger (HLT).

The whole Chapter 3 is devoted to the description of the trigger improvements that I developed during my Ph.D. One of the most important achievements is a fast tracking that allows to identify pile-up jets for any event accepted by the hardware trigger ($\sim 100\,\text{kHz}$). This fast tracking exploits a novel algorithm, the Fast Primary Vertex, that localizes the primary vertex without using tracks, but just hits recorded by the pixel detector and jets reconstructed by the calorimeters. The fast pile-up jet cleaning has been used to improve the resolution of the missing transverse energy and of the total transverse energy evaluated using calorimetric jets. The online b-tagging performance of CMS improved taking advantage of this new fast tracking and other developments. All the improvements have been used to develop new triggers dedicated to the search for the $H \to bb$ in the LHC Run 2, especially for $ZH \to \nu\nu bb$ channel. The rates of these triggers were estimated from simulations using a novel method that I developed, described in Appendix B. This method solved the problem of double counting of the rate due to pile-up collisions.

Chapter 4 is dedicated to the search for $H \to bb$ decay when the Higgs boson is produced in association with a Z boson decaying to neutrinos. The search has been performed using the $2.32\,\text{fb}^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ collected by CMS in 2015. This search is a preliminary result and will be updated with the new data delivered by LHC in 2016 ($\sim 20\,\text{fb}^{-1}$) and by combining it with the $WH \to \ell\nu$ and $ZH \to \ell\ell$ channels.

Contents

Chapter 1

The Standard Model Higgs boson

The Standard Model (SM) has precise predictions for all properties of the Higgs boson with the exception of its mass. This chapter, after introducing the SM of particle physics with a special attention to the Higgs sector, describes the properties of the SM Higgs boson with $m_H = 125\,\text{GeV}$ and compares them with the observed experimental results of the LHC Run 1.

1.1 The Standard Model

The SM of particle physics [1–3] describes with high precision all known phenomena in particle physics. A detailed description of SM can be found in [8–10]. In the SM, particles are grouped in two categories: particles with integer spin are called bosons and half-integer spin particles are called fermions. Among bosons, there are the gauge bosons which carry the fundamental interactions of nature and the Higgs boson that is related to the electroweak symmetry breaking. Among fermions, there are two classes of particles: quarks and leptons. Each class is divided into three generations, and each one is composed of a doublet of particles. The first generation of leptons is formed by the electron (e), with charge $Q = -1$ and mass $0.511\,\text{MeV}$, and the electronic neutrino (ν_e), that is neutral and almost massless [11]. The first generation of quarks is composed by the quark up (u), with $Q = +2/3$, and the quark down (d), with $Q = -1/3$. The masses of both particles are a few MeV. The other generations of particles are similar to the first one, but with higher mass. Quarks exist in three types, with different color charge: red (r), green (g), and blue (b).

Summarizing, in the SM there are three doublets of leptons:

$$\begin{pmatrix}\nu_e\\ e\end{pmatrix}\begin{pmatrix}\nu_\mu\\ \mu\end{pmatrix}\begin{pmatrix}\nu_\tau\\ \tau\end{pmatrix}, \tag{1.1}$$

and nine doublets of quarks:

$$\begin{pmatrix} u \\ d \end{pmatrix}_{r,g,b} \begin{pmatrix} c \\ s \end{pmatrix}_{r,g,b} \begin{pmatrix} t \\ b \end{pmatrix}_{r,g,b}. \tag{1.2}$$

For each fermion there is an antiparticle with opposite quantum numbers, same mass, and same spin.

1.1.1 Gauge theory

The kinematic of a free particle is defined by the action:

$$S = \int \mathcal{L}_{\text{free}} d^4x. \tag{1.3}$$

For fermions, $\mathcal{L}_{\text{free}}$ corresponds to the Dirac Lagrangian [12]:

$$\mathcal{L}_{\text{free}} = \bar{\psi}_i(i\gamma^\mu \partial_\mu - m_i)\psi_i \tag{1.4}$$

where ψ_i represent the fermionic fields, γ_μ are the Dirac matrices, and $\bar{\psi}_i = \psi_i^\dagger \gamma_0$.

If a symmetry holds for transformations depending on the space-time coordinates, it is called local symmetry.

Gauge theories [13] are quantum field theories where the Lagrangian has a local symmetry. Given the coupling constants, the symmetry group $\mathcal{G}$, and the transformations of the quantum fields under the group $\mathcal{G}$, gauge theories predict all the interactions between particles and the existence of the gauge bosons.

A generic local transformation of quantum field can be written as $\psi'^i(x) = U^{ij}(x)\psi^j(x)$, where $U^{ij}(x)$ depends on the space-time coordinates and is a linear combination of the generators of the symmetry group: $U^{ij}(x) = a^k(x)t_k^{ij}$. The space-time derivative of fields does not transform covariantly, i.e. like fields, and hence eq. (1.4) is not gauge invariant. To obtain a gauge-invariant Lagrangian, all space-time derivatives have to be replaced by the "covariant derivatives", defined as follows:

$$D_\mu \psi^i(x) = \partial_\mu \psi^i(x) - i\, g\, t_a^{ij}\, A_\mu^a(x)\psi^i(x) \tag{1.5}$$

where $A_\mu^a(x)$ are the gauge boson fields, t_a^{ij} are the generators of the group $\mathcal{G}$ in the representation of the fields ψ^i, and g is the coupling constant between ψ^i and $A_\mu^a(x)$. $D_\mu\psi^i(x)$ is covariant if and only if $A_\mu^a(x)$ transforms like:

$$A_\mu^a(x) \to U(x)\, A_\mu^a(x)\, U(x)^{-1} + \frac{i}{g}(\partial_\mu U(x))U(x)^{-1}. \tag{1.6}$$

The gauge-invariant kinetic term of the gauge bosons is given by:

$$\mathcal{L}_{\text{kin}} = -\frac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a, \tag{1.7}$$

where $F^a_{\mu\nu}$ is defined as:

$$F^a_{\mu\nu} = \frac{\left[D^a_\mu, D^a_\nu\right]}{ig} = \partial_\mu A^a_\nu(x) - \partial_\nu A^a_\mu(x) - igf^{abc} A^b_\mu(x) A^c_\nu(x). \tag{1.8}$$

The last term introduces trilinear and quadrilinear interactions among gauge bosons, in the non-abelian symmetries ($[T^a, T^b] = if^{abc}T^c \neq 0$). The symmetry forbids the presence of mass term for gauge bosons in the Lagrangian.

1.1.2 Electroweak unification

The SM is a gauge theory that contains three local symmetries:

$$SU(3)_C \times SU(2)_L \times U(1)_Y. \tag{1.9}$$

The strong nuclear interaction is described by $SU(3)_C$ and the electroweak interaction by $SU(2)_L \times U(1)_Y$. This chapter presents only the electroweak interaction since it is related to the Higgs boson physics.

Given the symmetry group $SU(2)_L \times U(1)_Y$, the field transformations are defined by the representation of the group. Under $SU(2)_L$, left-handed quarks and leptons transform like doublets in the fundamental representation ($T = \frac{1}{2}$). The doublets are formed as described in (1.1) and (1.2). Right-handed quarks and leptons are singlets under $SU(2)_L$ ($T = 0$). Each field has a quantum number, named weak isospin, that is $T_3 = 0$ for singlets and $T_3 = \pm\frac{1}{2}$ for doublets. The gauge bosons associated with the three generators of $SU(2)_L$ are called W_i. Under $U(1)_Y$, all fields are singlets and have an associated quantum number, called weak hypercharge, that is defined as $Y = Q - T_3$. Using this definition, the fields contained in each $SU(2)_L$ doublet have the same weak hypercharge. The gauge boson associated with the generator of $U(1)_Y$ is called B.

We can then write the covariant derivative of $SU(2)_L \times U(1)_Y$:

$$D_\mu = \partial_\mu - \frac{1}{2} i\, g\, W^i_\mu\, \sigma^i - i\, g'\, B_\mu, \tag{1.10}$$

where g e g' are the two coupling constants, and σ^i are the Pauli matrices. Applying the substitution $\partial_\mu \to D_\mu$, the Lagrangian becomes

$$\mathcal{L} = \bar{\psi}_i (i\gamma^\mu D_\mu - m_i)\psi_i - \frac{1}{4} F^a_{\mu\nu} F^{a\mu\nu}. \tag{1.11}$$

Still, it is has two problems: the gauge bosons are massless - contrary to observation - and the mass term $\bar{\psi}_i m_i \psi_i$ is not gauge invariant. Both problems are solved by the Higgs-Brout-Englert mechanism [4, 5].

1.1.3 Spontaneous symmetry breaking

Spontaneous symmetry breaking occurs when a Lagrangian is invariant under a symmetry, but the vacuum state breaks this symmetry. For example, let us consider the Lagrangian of a gauge theory for a complex scalar field φ:

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + (D_\mu \varphi)^* (D^\mu \varphi) + \mu^2 \varphi^* \varphi - \frac{\lambda}{4} (\varphi^* \varphi)^2 \tag{1.12}$$

where $D_\mu \varphi = \partial_\mu \varphi - i\, e\, A_\mu\, \varphi$. In this example, the Lagrangian is $U(1)$ gauge invariant and, if $\mu^2 < 0$, it describes the kinematic of a particle having mass $\sqrt{-\mu^2}$. In case $\mu^2 > 0$, the minimum of the potential is no longer in $\varphi = 0$, but it is in the circumference $|\varphi|^2 = \frac{2\mu^2}{\lambda} = v^2$. There are an infinite number of possible potential minima (vacuum states) and each minimum is no longer $U(1)$ invariant.

Parameterizing the field $\varphi(x)$ as $\varphi(x) = \rho(x) e^{i\,\theta(x)}$, and expanding around v, the covariant derivative becomes

$$D_\mu \varphi(x) = [\partial_\mu \rho(x) - i\, \rho(x)(\partial_\mu \theta(x) - e\, A(x)_\mu) - i\, e\, A(x)_\mu\, (\rho(x) - v)]\, e^{i\,\theta(x)}. \tag{1.13}$$

A gauge transformation of parameter Λ transforms fields as $\theta(x) \to \theta(x) + e\Lambda$ and $A_\mu(x) \to A_\mu(x) - \partial_\mu \Lambda$. Hence, we can choose $\Lambda(x)$ to eliminate $\theta(x)$ from the $|D_\mu \varphi|^2$ term in the Lagrangian:

$$|D_\mu \varphi(x)|^2 = (\partial_\mu \rho(x))^2 + e^2 (\rho(x) - v)^2 A(x)_\mu A(x)^\mu. \tag{1.14}$$

Using this parametrization, the Lagrangian does not depend on $\theta(x)$ and the gauge boson is massive since $e^2\, v^2\, A(x)_\mu\, A(x)^\mu$ is a mass term. The Lagrangian contains also trilinear and quadrilinear vertices ρA^2 and $\rho^2 A^2$, which give the interactions between the gauge boson and the scalar field. In addition, the expansion around $\rho = v$ of the potential $V(\varphi) = -\mu^2 \varphi^* \varphi + \frac{\lambda}{4} (\varphi^* \varphi)^2 = -\mu^2 \rho^2 + \frac{\lambda}{4} \rho^4$ gives both a mass term (ρ^2) and self-coupling terms (ρ^3, ρ^4) for the scalar field ρ.

This mechanism of giving mass to the gauge bosons through the spontaneous symmetry breaking is generic and valid for other symmetry groups too. It was conceived by P. Higgs [4], R. Brout and F. Englert [5] in 1964.

1.1.4 Mass problem for gauge bosons

The Higgs-Brout-Englert mechanism is used in the SM to give mass to the electroweak gauge bosons ($W^{\pm}$, Z^0). This is achieved by adding two new complex scalar fields which form a $SU(2)_L$ doublet: $\Phi(x) = \binom{\varphi^+(x)}{\varphi(x)}$. The potential of the new field $\Phi(x)$ has a minimum in $\Phi(x) = \binom{0}{v}$, and this breaks spontaneously the $SU(2)_L \times U(1)_Y$ symmetry.

The field $\Phi(x)$ can be expressed as a function of v and $\theta(x)^a$:

$$\Phi(x) = e^{\frac{1}{2}\, i\, \theta(x)^a \sigma^a} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix}. \tag{1.15}$$

After the spontaneous symmetry breaking, the Lagrangian no longer depends on the fields $\theta(x)^a$ and contains the mass terms for the gauge bosons:

$$|D_\mu \Phi|^2 = \frac{1}{4}\left[g^2(|W^1|^2 + |W^2|^2) + (g\, W_\mu^3 - g'\, B_\mu)^2\right]\left[v^2 + \frac{h^2(x)}{2} + \frac{2v\, h(x)}{\sqrt{2}}\right]. \tag{1.16}$$

Finally, we can rotate the field (W^1, W^2) and (W^3, B) to find the mass eigenstates of the gauge bosons:

$$\begin{aligned} W_\mu^{\pm} &= \tfrac{1}{\sqrt{2}}(W_\mu^1 \pm i\, W_\mu^2) \\ Z_\mu^0 &= \tfrac{1}{\sqrt{g^2+g'^2}}(g\, W_\mu^3 - g'\, B_\mu) \\ A_\mu &= \tfrac{1}{\sqrt{g^2+g'^2}}(g\, W_\mu^3 + g'\, B_\mu) \end{aligned} \tag{1.17}$$

and the mass eigenvalues: $m_{W^\pm} = g\frac{v}{2}$, $m_Z = \sqrt{g^2 + g'^2}\frac{v}{2}$, $m_A = 0$.

Note that Eq. (1.16) predicts also the interaction between the Higgs field $h(x)$ and the massive gauge bosons Z^0 and $W^{\pm}$.

1.1.5 Mass problem for quarks and leptons

The fermionic mass term $m\bar{\psi}^i\psi^i$ is not $SU(2)_L$ invariant. This problem is solved by adding a new interaction between the Higgs field and fermions. The electron mass can be introduced by adding to the Lagrangian the following term:

$$\begin{aligned} & -G_e\left[\bar{\psi}_{e,L}\Phi\psi_{e,R} + \bar{\psi}_{e,R}\bar{\Phi}\psi_{e,L}\right] \\ = \;& -G_e\left[(\bar{\nu_e}, \bar{e})_L \cdot \binom{\varphi^+}{\varphi^0} e_R + \bar{e_R}(\varphi^-, \bar{\varphi^0}) \cdot \binom{\nu_e}{e}_L\right] \\ = \;& -\tfrac{G_e}{\sqrt{2}} v\left[\bar{e_L} e_R + \bar{e_R} e_L\right] - \tfrac{G_e}{\sqrt{2}}\left[\bar{e_L} e_R + \bar{e_R} e_L\right] h(x) \\ = \;& -m_e\left[\bar{e}e\right] - \tfrac{m_e}{v} h(x)\left[\bar{e}e\right] \end{aligned} \tag{1.18}$$

where $G_e = \sqrt{2} m_e / v$. Likewise, we can add the mass term for the other charged leptons and up-type quarks. Down-type quark masses can be obtained by a similar method using the doublet $\Phi^C = -i\,\sigma^2\,\Phi^*$ that has the expectation value $\binom{v}{0}$. In addition, the down-type quark mass eigenstates are rotated with respect to the $SU(2)_L$ eigenstates

by the Cabibbo-Kobayashi-Maskawa matrix [14, 15]. Therefore, the quark mass term in the Lagrangian appears as follows:

$$\begin{aligned} & -G_d^{ij}(\bar{u_i}, \bar{d_i'})_L \cdot \begin{pmatrix} \varphi^+ \\ \varphi^0 \end{pmatrix} d_{jR} - G_u^{ij}(\bar{u_i}, \bar{d_i'})_L \cdot \begin{pmatrix} \bar{-\varphi^0} \\ \varphi^- \end{pmatrix} u_{jR} + h.c. \\ = \; & -m_d^i \bar{d}_i d_i (1 + \tfrac{h(x)}{v}) - m_u^i \bar{u}_i u_i (1 + \tfrac{h(x)}{v}) \end{aligned} \tag{1.19}$$

Note that both Eq. (1.18) and (1.19) predict the interactions between the Higgs field and the fermions.

1.2 The Higgs boson at the hadron colliders

1.2.1 Production and decay channels

In the previous section, it has been shown that the Higgs boson couples with all massive particles, in particular to fermions and to the vector bosons Z and W. As gluons and photons are massless, they cannot interact directly with the Higgs boson, however they can couple to the Higgs boson through virtual loops, that are dominated by the top-quark loop.

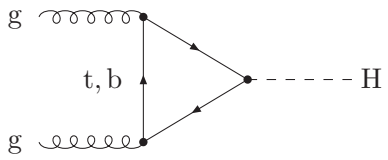
At the LHC, the main Higgs boson production channels are four: the gluon-gluon fusion (ggH), the vector-boson fusion (VBF), the vector-boson associated production (VH), and the top-pair associated production (ttH). The cross sections of the production channels at $\sqrt{s} = 13\,\text{TeV}$ are reported in Table 1.1, and they are in the range from 0.5 pb to 40 pb for a Higgs mass of 125 GeV.

The signal topologies in the four production channels are different:

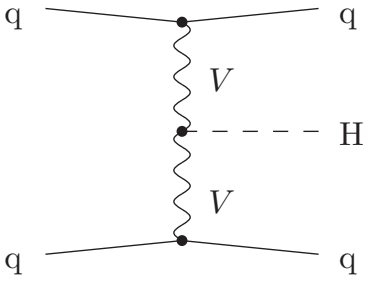
1. in ggH, Figure 1.1a, the only objects which are expected to have high transverse momentum are the Higgs decay products;
2. in VBF, Figure 1.1b, two forward/backward jets produced by the hadronization of the interacting proton quarks are expected, in addition to the Higgs boson. No extra hadrons are produced in the angular region between the two jets;

TABLE 1.1: SM Higgs boson production cross section at $m_H = 125.0\,\text{GeV}$ and $\sqrt{s} = 13$ TeV [16].

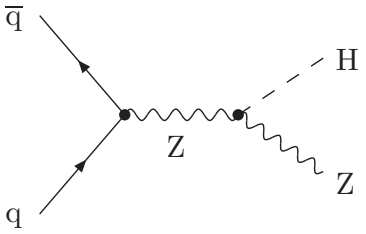
Process	Cross section (pb)
Gluon-gluon fusion	43.9 ± 4.4
Vector-boson fusion	3.75 ± 0.12
Vector-boson associated production	2.25 ± 0.06
Top-pair associated production	0.51 ± 0.05



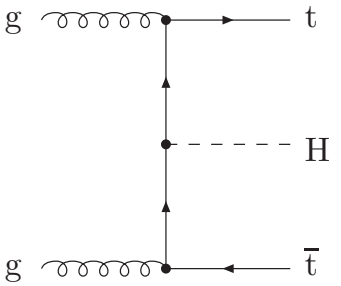
(A) Gluon-gluon fusion.



(B) Vector-boson fusion.



(C) Vector-boson associated production.



(D) Top-pair associated production.

FIGURE 1.1: Feynman diagrams of the Higgs boson production channels [17].

3. in VH, Figure 1.1c, the signal topology is characterized by the presence of a Higgs boson and a vector boson that are back-to-back on the transverse plane;

4. in ttH, Figure 1.1d, two top quarks are produced in association with the Higgs boson.

Table 1.2 shows the Higgs boson branching ratios expected for $m_H = 125\,\text{GeV}$.

1.2.2 Experimental results after the LHC Run 1

During the first run of LHC [18], the CMS [6] and ATLAS [7] experiments have collected an integrated luminosity of proton-proton collision of about $5\,\text{fb}^{-1}$ each at a center-of-mass energy of $\sqrt{s} = 7\,\text{TeV}$, and $20\,\text{fb}^{-1}$ at $\sqrt{s} = 8\,\text{TeV}$.

All direct and indirect measurements on the Higgs boson properties are compatible with the SM Higgs boson. In particular, the couplings of the Higgs boson to the Z and W bosons are determined with an experimental accuracy of $\sim 10\%$ and to the top quark with an accuracy of $\sim 15\%$. These measurements have been done observing the decays $H \to \gamma\gamma$ [19, 20], $H \to ZZ^{(*)} \to 4\,\ell$ [21, 22], $H \to WW^{(*)} \to 2\ell 2\nu$ [23, 24] ($\ell = \mu, e$), each one with more than five standard deviations. The search for the $H \to \gamma\gamma$ is characterized by a narrow peak in the diphoton invariant mass distribution above a smooth background, low signal-to-background ratio $O(0.1)$, and moderate signal yield;

the $H \to ZZ^{(*)} \to 4\ell$ by very high signal-to-background ratio $O(> 1)$ and small signal yield; the $H \to WW^{(*)} \to 2\ell 2\nu$ has a large signal yield, but a low signal-to-background ratio $O(0.1)$, due to the poor invariant mass resolution caused by the presence of the neutrinos and to the large irreducible $pp \to WW$ background.

The Higgs-τ coupling is measured with a larger uncertainty, of about 20%, due to the lower sensitivity of the $H \to \tau\tau$ analysis [25, 26]. Indeed, in this channel the presence of neutrinos degrades the resolution on the Higgs invariant mass, making $Z \to \tau\tau$ the main background. In the $H \to \tau\tau$ decay channel, an excess of 3 standard deviations has been observed.

The Higgs-b coupling [27–31] has the largest uncertainty ($\sim$ 25%) essentially for two reasons:

- the QCD multijet background $pp \to bb + X$, that overwhelms the gluon-gluon fusion $H \to bb$ process and reduces the sensitivity of the VBF $H \to bb$;
- the poorer mass resolution of the di-jet channel compared to the $\gamma\gamma$ or four leptons channels.

The invisible Higgs decay can be observed as missing transverse energy in the VBF or VH production channels. Measurements performed by CMS and ATLAS exclude a branching ratio $H \to$ invisible larger than 0.3 [32, 33].

A combination of the searches for the SM Higgs boson performed by the ATLAS and CMS collaborations after Run 1 is reported in [34]. Figure 1.2 shows the signal strength

TABLE 1.2: SM Higgs boson branching ratios for $m_H = 125.0\,\text{GeV}$. The branching ratio of the Higgs boson decaying to vector bosons with leptons in final state ($\ell = \mu, e$) are reported in the bottom of the table [16].

Decay channel	Branching ratio (%)
$H \to b\bar{b}$	57.7 ± 1.8
$H \to WW^*$	21.5 ± 0.9
$H \to gg$	8.57 ± 0.87
$H \to \tau\bar{\tau}$	6.32 ± 0.36
$H \to c\bar{c}$	2.91 ± 0.35
$H \to ZZ^*$	2.64 ± 0.11
$H \to Z\gamma$	0.154 ± 0.013
$H \to \gamma\gamma$	0.228 ± 0.011
$H \to s\bar{s}$	0.025 ± 0.001
$H \to \mu\bar{\mu}$	0.022 ± 0.001
$H \to WW^* \to 2\ell 2\nu$	0.977 ± 0.041
$H \to ZZ^* \to 4\ell$	0.0119 ± 0.0005

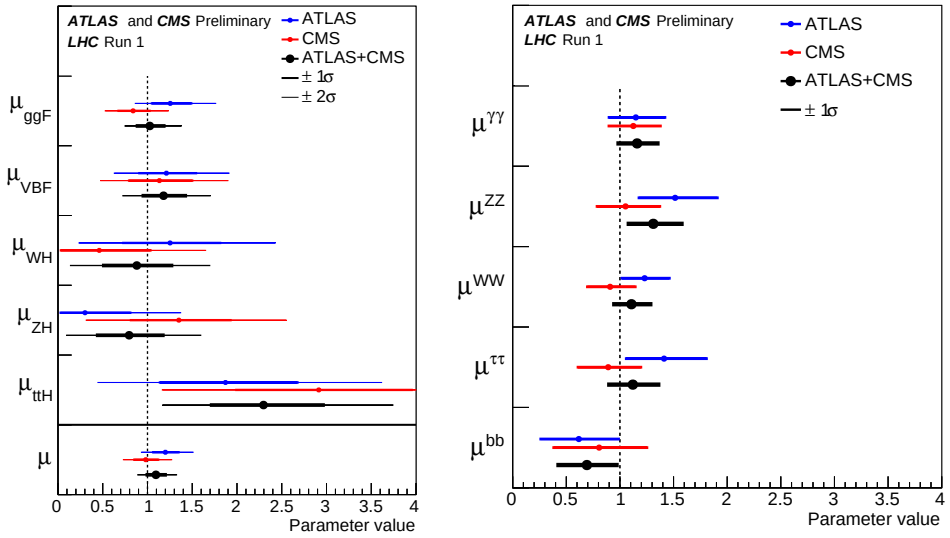


FIGURE 1.2: Combinations of the Higgs boson signal strengths measured by the ATLAS and CMS collaborations using the whole Run-1 dataset. On plot (A), the signal strengths are measured for the five main production channels. On plot (B), the signal strengths are shown for the different decay channels. The plots report the values measured by the ATLAS (blue) and CMS (red) collaborations, and the combination of them (black) [34].

$\mu = \sigma/\sigma_{SM}$ measured for the different production and decay channels of the Higgs boson. All measurements are compatible with the SM predictions within two standard deviations, but in the ttH channel both ATLAS and CMS have reported an excess larger than one and two standard deviations, respectively.

Figure 1.3 shows the ratios of the branching ratios and of the cross sections of the production channels with respect to the $gg \to H \to ZZ$ process, normalized to the SM predictions. A large part of the measurements have been found to be in agreement with the SM expectation within two standard deviations. However, the production cross-section ratio $\sigma_{\mathrm{ttH}}/\sigma_{\mathrm{ggF}}$ relative to the SM ratio has been measured to be $3.3^{+1.0}_{-1.4}$, corresponding to an excess of about 2.3 standard deviations, and the ratio $\mathrm{BR}^{bb}/\mathrm{BR}^{ZZ}$ relative to the SM ratio has been measured to be $0.19^{+0.21}_{-0.12}$. Figure 1.4 shows the likelihood scan of $\mathrm{BR}^{bb}/\mathrm{BR}^{ZZ}$ and the deficit reported with respect to the SM expectation is of approximately 2.5 standard deviations. The ttH excess is mainly due to the multi-leptons categories but no excess have been found in the $ttH(bb)$ analysis. This makes the $\mathrm{BR}^{bb}/\mathrm{BR}^{ZZ}$ anti-correlated with the $\sigma_{\mathrm{ttH}}/\sigma_{\mathrm{ggF}}$ and it explains why the fitted value for $\mathrm{BR}^{bb}/\mathrm{BR}^{ZZ}$ is so small.

Finally, the searches for the Higgs boson have been combined to fit the couplings of the fundamental particles with the Higgs field. The results are shown in Fig. 1.5. No significant deviations from the SM predictions are reported and the dependence of the couplings from the particle masses has been found in good agreement with the SM prediction.

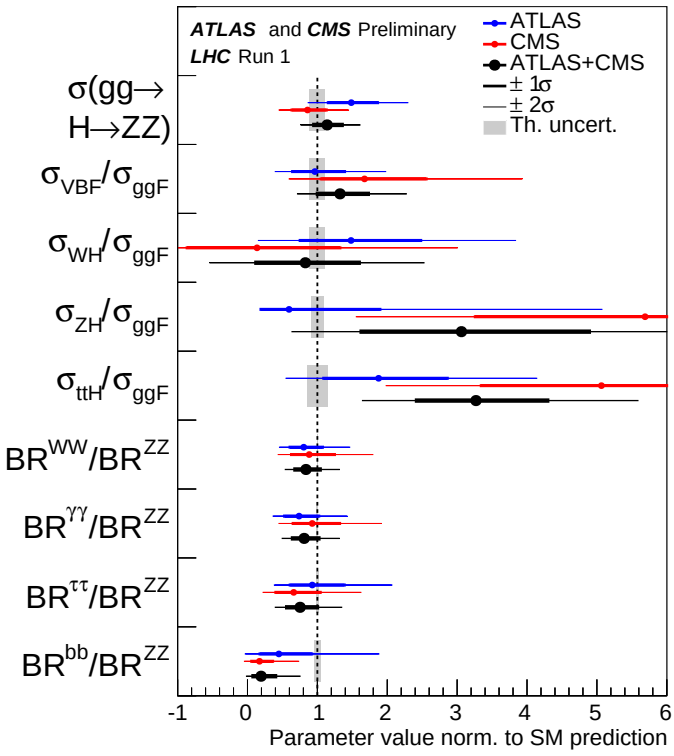


FIGURE 1.3: Fitted value of the $\sigma(gg \to H \to ZZ)$ cross section and ratios of cross-section and branching ratios with respect to the $\sigma(gg \to H \to ZZ)$. The values have been normalized to the SM prediction. The plot reports the values measured by the ATLAS (blue) and CMS (red) collaborations, and the combination of them (black) [34].

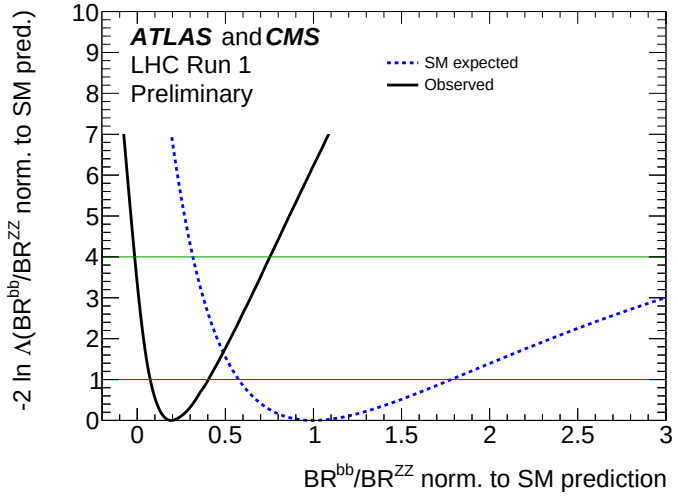
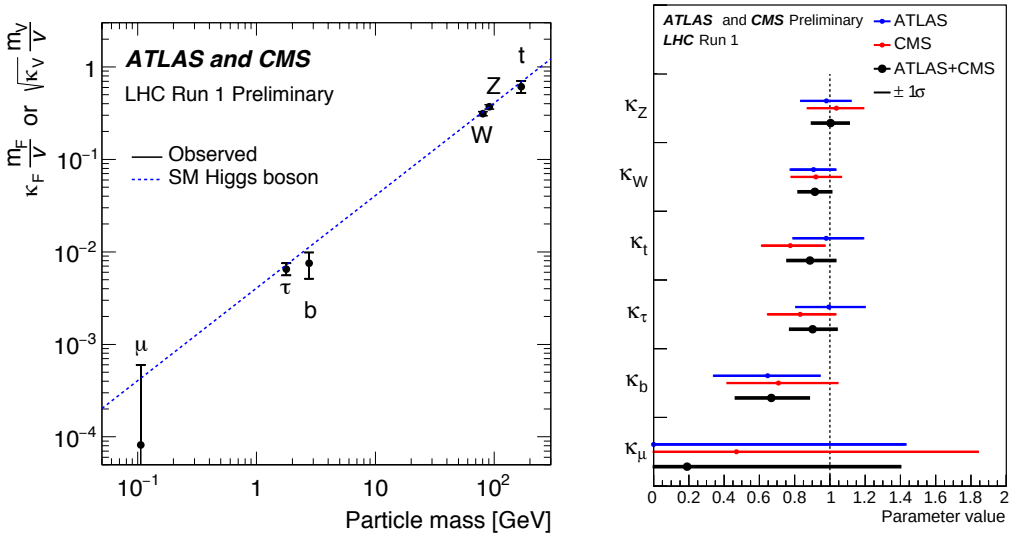


FIGURE 1.4: Observed (black line) and expected (blue line) negative log-likelihood scan of the $\mathrm{BR}^{bb}/\mathrm{BR}^{ZZ}$ parameter [34].



(A) (B)

FIGURE 1.5: Combinations of the Higgs boson couplings measured by the ATLAS and CMS collaborations using the whole Run-1 dataset. On plot (A), the Yukawa coupling, for fermions, and square root of the coupling, for the vector bosons, are reported as a function of the mass. On plot (B), the ratio of the couplings measured by the ATLAS (blue), CMS (red), and ATLAS+CMS (black) collaborations over the SM expectations are reported [34].

Chapter 2

The CMS experiment at the LHC

After a description of the CMS experiment at the LHC, this chapter describes how the proton-proton collisions are reconstructed with the CMS experiment and, specifically, how the main $Z(\nu\nu)H(bb)$ objects (jet, missing transverse energy, and b tagging) are identified. The last section of this chapter presents the trigger system of CMS, with special attention to the High Level Trigger where I gave my most important contributions.

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [18] is a circular proton-proton collider built close to Geneva by the European Organization for Nuclear Research (CERN). It is located in a 27 km long tunnel that previously hosted the Large Electron-Positron collider (LEP). It was designed to collide protons at the center-of-mass energy of $\sqrt{s} = 14\,\text{TeV}$ at the instantaneous luminosity of $\mathcal{L} = 1 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. The luminosity delivered by the LHC during the first run (2010-2012) was about $6\,\text{fb}^{-1}$ and $23\,\text{fb}^{-1}$ at 7 TeV and 8 TeV, respectively. In 2015, after a two years stop, the LHC started Run 2 delivering the first proton collisions at $\sqrt{s} = 13\,\text{TeV}$. LHC reached a luminosity of $5 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ and integrated about $4\,\text{fb}^{-1}$, as shown in Fig. 2.1. In the next years, the LHC is expected to reach a luminosity of about $2 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, reaching $\sim 100\,\text{fb}^{-1}$ of integrated luminosity by 2018.

In the LHC, protons circulate in bunches of about 5 cm-length. During Run 2, the separation distance between bunches has been decreased from $\sim 15\,\text{m}\,(50\,\text{ns} \cdot \text{c})$ to $\sim 7.5\,\text{m}\,(25\,\text{ns} \cdot \text{c})$. When two bunches cross each other, there are simultaneous proton-proton collisions. The mean number of simultaneous collisions is determined by the following equation:

$$\text{collisions per bunch crossing} = f \cdot d/c = \mathcal{L} \cdot \sigma_{pp} \cdot d/c \tag{2.1}$$

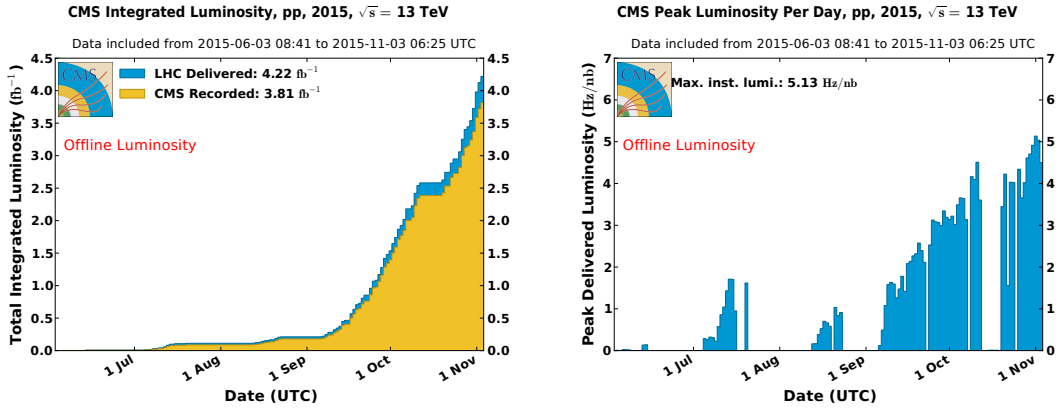


FIGURE 2.1: In blue, the integrated (left) and peak instantaneous (right) luminosity delivered by LHC to the CMS experiment in 2015. In yellow on the left, the integrated luminosity recorded by the CMS experiment [35].

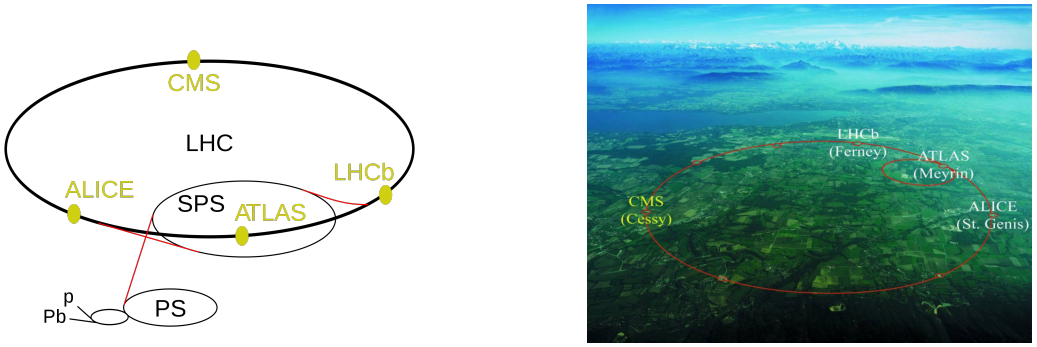
where f is the frequency of the proton-proton collisions, d is the mean separation distance among bunches, c is the speed of the protons, $\mathcal{L}$ is the instantaneous luminosity, and $\sigma_{pp} \sim 70\,\text{mb}$ is the proton-proton cross section at 13 TeV. The mean number of simultaneous collisions was about 22 in Run 1 and 18 in 2015. In the next few years, when LHC will reach $\mathcal{L} \sim 2 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, up to 50 simultaneous proton-proton collisions are expected.

Protons are confined in a circular orbit by a magnetic field of $B = 8.3\,\text{T}$ produced by superconducting dipoles. The magnets are maintained at $T = 1.9\,\text{K}$ by a cooling system based on superfluid helium. Before being delivered to LHC, protons are accelerated through a complex of linear and circular accelerators, as shown by fig. 2.2a. A detailed description of LHC is available in [18].

At LHC, there are four main experiments, as shown in Fig. 2.2: A Toroidal LHC ApparatuS (ATLAS) [7], Compact Muon Solenoid (CMS) [6], Large Hadron Collider beauty (LHCb) [36], and A Large Ion Collider Experiment (ALICE) [37]. CMS and ATLAS are two multi-purpose experiments: they have been designed to investigate many fields of particle physics, including precision measurements of SM processes, search for the Higgs boson, and search for supersymmetry and extra dimensions. The main goal of the LHCb experiment is the precision measurement of processes related to the b quark and the CP-symmetry violation. LHC is also able to deliver heavy ions collisions and ALICE is an experiment dedicated to the study of these collisions.

2.2 The CMS detectors

The Compact Muon Solenoid (CMS), Figure 2.3, is a multipurpose experiment designed to measure the particles originated by the hadron-hadron collisions provided by the



(A) (B)

FIGURE 2.2: The accelerator complex and the four experiments at the LHC [38].

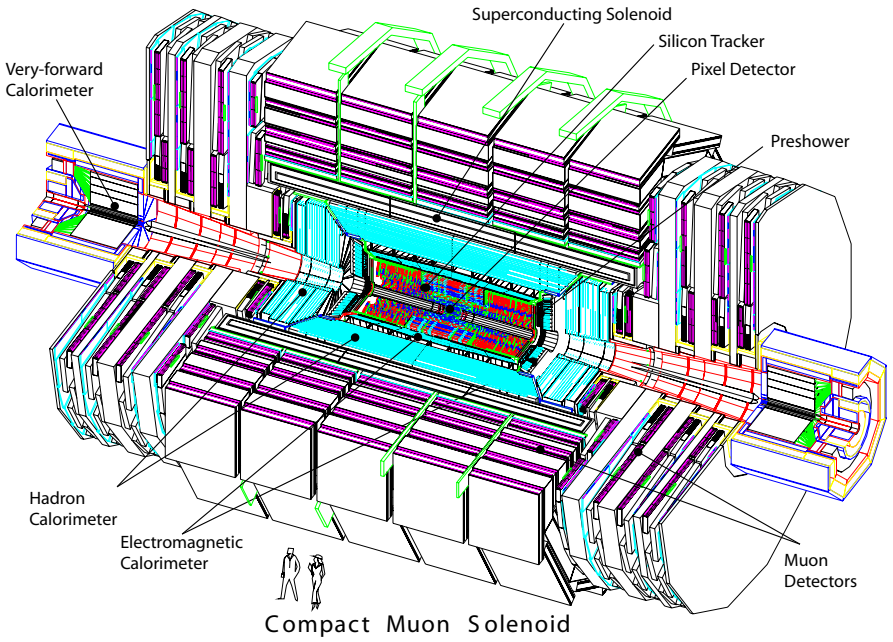


FIGURE 2.3: The CMS experiment [39].

LHC. This section provides a brief description of the CMS subdetectors. A complete description of the detector can be found in [39].

The coordinate system adopted by CMS is a right-handed set of axises with the origin in the nominal collision point. The x-axis is pointing towards the center of LHC and the y-axis is directed upward perpendicular to the plane of the ring, therefore the z-axis is parallel to the beam line and it is pointing towards the Jura mountains. The azimuthal angle φ is measured from the x-axis on the $x - y$ plane, and the radial coordinate is named r. The polar angle θ is measured from the z-axis. The pseudorapidity is defined as $\eta = -\ln(\tan\theta/2)$. The angular cone $\Delta R = \sqrt{\Delta\eta^2 + \Delta\varphi^2}$ is used to identify angular regions of interest for many applications like jet definition and isolation.

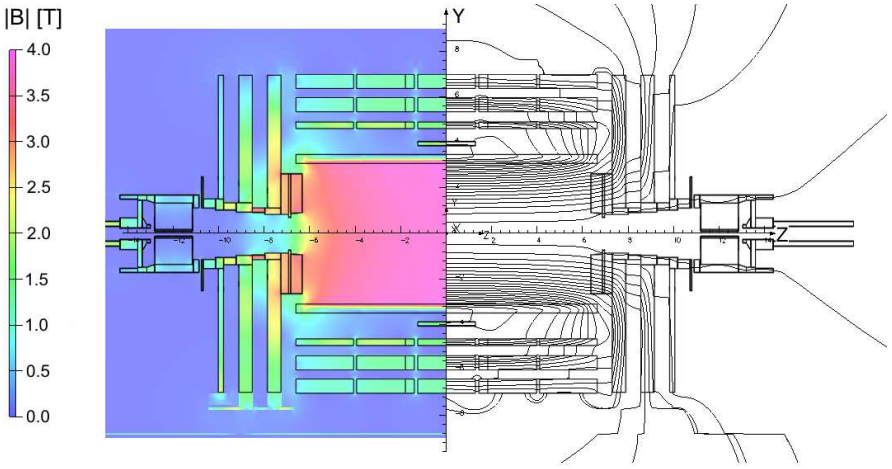


FIGURE 2.4: Map of the $|B|$ field (left) and field lines (right) expected for a longitudinal section of the CMS experiment [40].

2.2.1 The superconducting solenoid

At the CMS, both the tracker and the calorimeters are immersed in an approximately uniform magnetic field produced by a superconducting solenoid. The magnetic field is about 3.8 T and is directed towards the z-axis. The magnetic flux generated by the current in the solenoid is closed with several instrumented iron structures as shown in Fig. 2.4.

The superconducting magnet is kept at about 4 K by a cooling system based on liquid helium. In 2015, about a fifth of the integrated luminosity has been collected without magnetic field because of problems related to the cryogenic system.

2.2.2 The silicon tracker

The silicon tracker is the innermost detector of CMS. It has been designed to detect the trajectories of charged particles up to $|\eta| \approx 2.5$. The tracker is composed by approximately 16000 modules arranged as shown in Fig. 2.5. The surface of the sensors increases with the radius. Innermost modules have the electrodes in pixel shape and the outer modules have the electrodes in strips, measuring with precision the coordinate perpendicular to the magnetic field ($r-\varphi$). The sensors are composed of a n-bulk device with a single p^+-doped face, for strip modules, and of a p-doped silicon wafer, and two faces p^+- and n^+-doped, for pixel modules. Figure 2.6 is a sketch of the sensors of a pixel module. In each module, a wide depleted volume is obtained with a strong electric field of $\sim 1000\,\mathrm{V/mm}$. When charged particles pass through this region, electron-hole pairs are created and the two partners are drifted in opposite directions by the electric field. As electrons are faster than holes, they induce the strongest current signal. In the barrel modules, where the electric field is perpendicular to the magnetic field, the

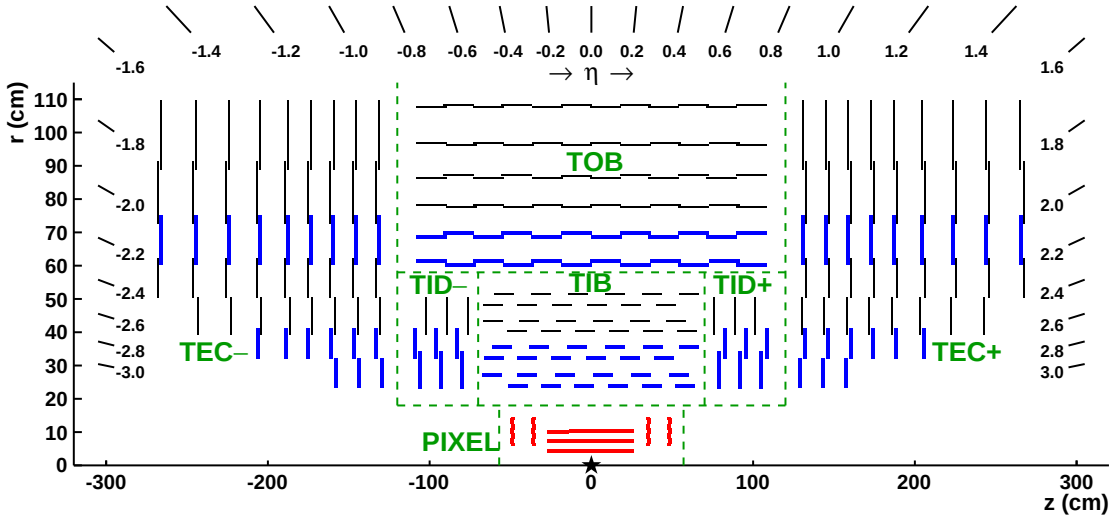


FIGURE 2.5: Longitudinal section of the silicon tracker. In details: the red lines are the pixel modules, the thin lines are the single strip modules, and the blue lines are the double strip modules [41].

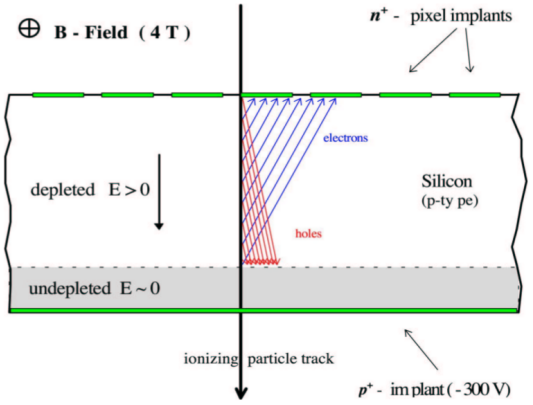


FIGURE 2.6: A pixel module is set up by three components: a p-doped silicon wafer, a heavy p-doped side, and n-doped pixels on the other side. The motion of electrons and holes generated by ionizing particles is given by the electric and magnetic field. The electrons are directed towards the n-doped side because an inverted bias voltage is applied. The magnetic field, orthogonal to the electric field in the pixel barrels, deviates the motion by the Lorentz angle defined in (2.2) [42].

drift velocity of the electrons towards the n-doped side has an angle with respect to the direction of the electric field, the so-called Lorentz angle. This angle is equal to:

$$\tan\theta_\ell = \frac{F_x}{F_z} = \frac{v_z B}{E} = \mu B \implies \theta_\ell \sim 20^\circ(\text{pixel}), 5^\circ(\text{strip}) \tag{2.2}$$

where μ is the electron mobility in the module. In order to improve the spatial resolution, the pixel tracker has been designed so that high p_T tracks signal is spread in at least two sensors on the transverse plane.

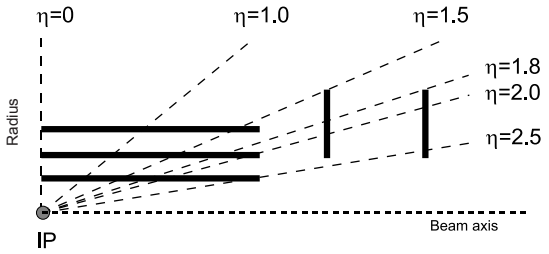


FIGURE 2.7: The pixel tracker on the longitudinal plane [39].

The pixels

The pixels are the detectors placed closest to the nominal collision point ($4.4\,\text{cm} < r < 10.2\,\text{cm}$) and, therefore, they have a key role in reconstructing tracks and vertices. These modules are exposed to a strong radiation produced by the proton collisions (up to 10^8 particles $\text{cm}^{-2}\text{s}^{-1}$). The pixel modules are arranged in three cylindrical barrels and two end-cap disks, as shown in Figure 2.7. In order to reduce the occupancy and improve the position resolution, pixel sensors have been designed with a small area ($100 \times 150\,\mu\text{m}^2$ on $\varphi - z$ plane) and a thickness of $285\,\mu\text{m}$.

The strips

The strip modules are positioned externally to the pixel detectors ($0.2\,\text{m} < r < 1.1\,\text{m}$). The strip sensors have a rectangular shape with a width of about $10\,\mu\text{m}$ and a length of $10 - 20\,\text{cm}$, depending on their distance from the beam line. The pitch between two sensors is $60 - 200\,\mu\text{m}$. Strips have been designed to have high resolution on the position on the transversal plane to the magnetic field for a precise measurement of the track momentum.

In order to measure also the longitudinal coordinate, some regions have been equipped with pairs of strip modules, forming an angle of $\sim 100\,\text{mrad}$ between them. This small angle was chosen to both measure the longitudinal coordinate with a good resolution and to reduce the problem of the hit ambiguity.

2.2.3 The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) measures the energy and the direction of photons and electrons. ECAL is a homogeneous calorimeter composed of 61200 lead-tungstate crystals ($PbWO_4$) in the barrel e 7324 crystals in each end-cap. The crystals cover up to $|\eta| \approx 3.0$ and their shape is a truncated pyramid. The ECAL granularity is very high: the angular size of each barrel crystal is $\Delta\eta \times \Delta\varphi = 0.0174 \times 0.0174$.

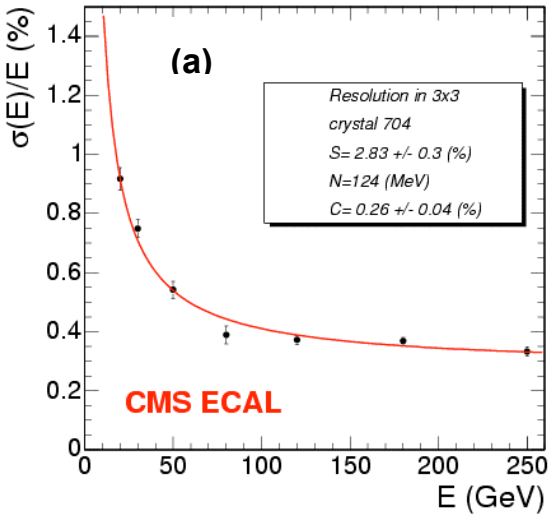


FIGURE 2.8: The energy resolution of ECAL for electrons as a function of energy [44].

When photons or electrons hit a crystal they produce an electromagnetic shower. The $PbWO_4$ is a scintillating material and a fraction of the ionization energy released by the charged particles in the shower is collected in a visible light signal. The light is measured by avalanche photodiodes (APD) in the barrel and vacuum phototriodes (VPT) in the endcaps. Thanks to the very small S/N ratio, the crystals are also sensitive to the primary scintillation of non-showering particles as muons.

The energy resolution of the calorimeter was measured with an electron beam. The results are shown in Fig. 2.8. The resolution of the electron energy measurement is:

$$\left(\frac{\sigma(E)}{E}\right)^2 = \left(\frac{2.8\%}{\sqrt{E}}\right)^2 + \left(\frac{12\%}{E}\right)^2 + (0.30\%)^2 \tag{2.3}$$

where E is in GeV [43].

The $PbWO_4$ was chosen as radiator because of the small radiation length (needed to have a compact calorimeter with good energy resolution), the fast response (80% of the signal is collected in ≤ 25 ns), the radiation and magnetic hardness, and the small Molière radius that gives a high angular resolution.

In the endcaps, the internal face of the calorimeter is equipped with a preshower. It is composed of two silicon detectors planes with a 1 cm lead plate in the middle. The preshower is useful to discriminate the isolated photon or electron signals from the $\pi_0 \to \gamma\gamma$ background. This detector helps in discriminating between showers induced by hadrons and electrons/photons.

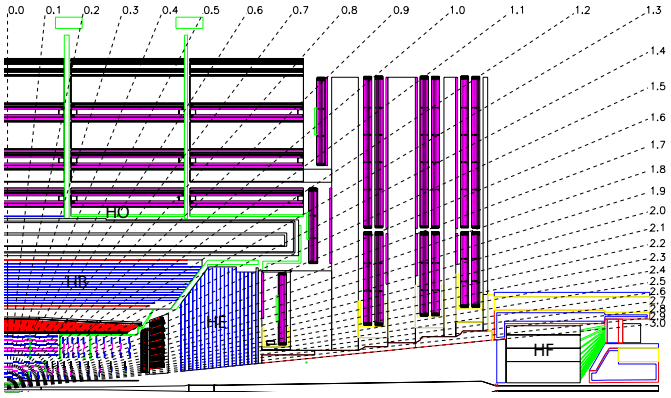


FIGURE 2.9: The longitudinal section of the hadron calorimeter of CMS [6].

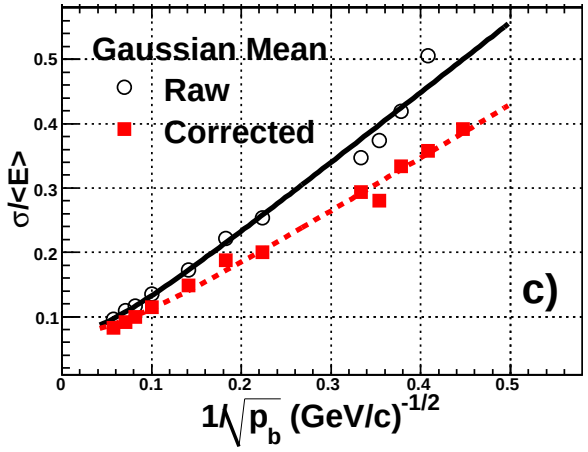


FIGURE 2.10: The energy resolution of HCAL-HB for a π^+ beam. The resolution has been evaluated before (circles) and after (squares) a set of preliminary corrections [45].

2.2.4 The hadron calorimeter

The hadron calorimeter (HCAL) is the detector used to measure the energy of hadrons, in particular of the neutral ones. HCAL is a heterogeneous calorimeter: the absorber is brass and the energy is measured through plastic scintillators crossed by the particles of the hadronic shower.

The visible-light signal is collected with fibers and is measured by hybrid photodiodes (HPD). As shown in Fig. 2.9, HCAL is made of three components in $|\eta| < 3$: the barrel (HCAL-HB), the end-caps (HCAL-HE), and the external calorimeter (HCAL-HO). The angular segmentation of HCAL is $\Delta\eta \times \Delta\varphi = 0.087 \times 0.087$. Figure 2.10 shows the energy resolution of HCAL for a beam of π^+.

The detector acceptance is extended with the two forward calorimeters (HCAL-HF) (from $|\eta| \approx 3$ to $|\eta| \approx 5$). They are placed outside the magnetic field, and they use iron as an absorber and the Cherenkov light in quartz fibers to measure the shower energy. The light signal is measured by photomultipliers (PMT).

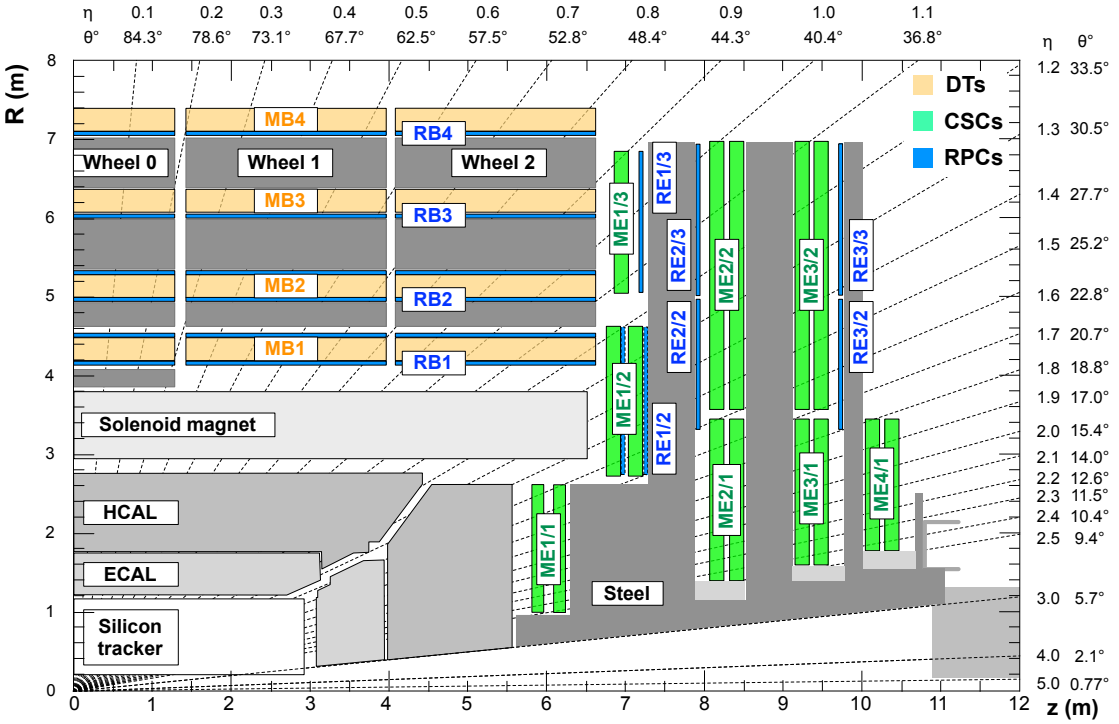


FIGURE 2.11: The CMS muon chambers [6].

2.2.5 The muon chambers

The muon chambers are the outermost detectors in CMS. They are used to detect muons which, together with neutrinos, are the only known particles escaping from CMS. Three kinds of detectors are used in the muon chambers: Drift Tube (DT), Cathode Strip Chamber (CSC), and Resistive Plate Chamber (RPC). They are placed in the steel return jokes of the magnetic field, as shown in Fig. 2.11, where the magnetic field strength is around 1 – 2 T, as shown in Fig. 2.4.

Figure 2.12 shows the number of interactions lengths between the collision point and the CMS subdetectors. Note that, for HCAL, there are at least 7 interaction lengths, and the probability for a hadron to pass through HCAL without interacting is very low: $< e^{-7} \approx 9 \cdot 10^{-4}$. Most important is the negligible probability that a hadron crosses the whole CMS detector without interacting: $\lesssim e^{-20} \approx 10^{-9}$.

2.3 The event reconstruction

This section describes how tracks, vertices, electrons, photons, jets, missing transverse energy, and muons are reconstructed in the CMS experiment. This section introduces also the main algorithms used for b tagging. The particle flow (PF) algorithm [46] combines the information of all CMS subdetectors in order to improve the reconstruction performance.

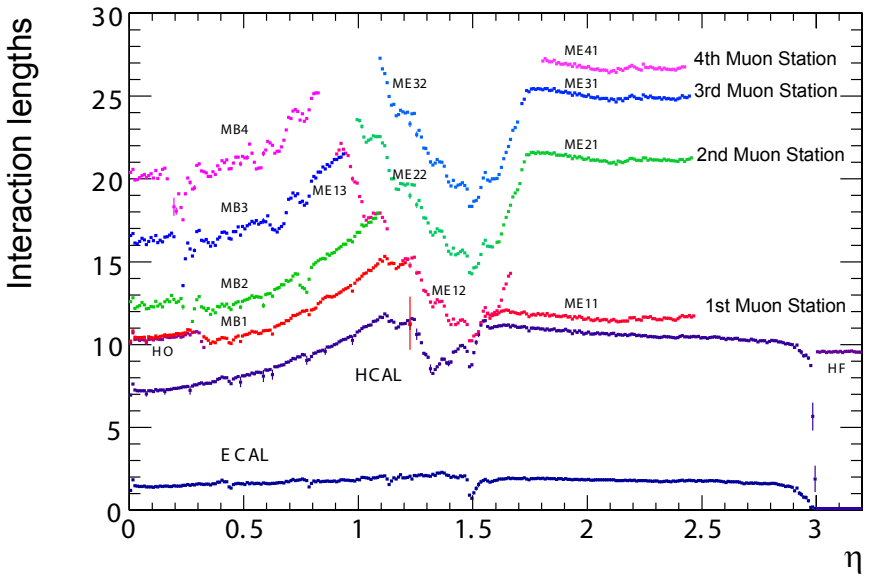


FIGURE 2.12: Interaction lengths, as a function of η, between the collision point and ECAL, HCAL, and the four muon stations [6].

2.3.1 The tracks

Tracks are the reconstructed trajectories of charged particles. They serve many purposes like finding the primary vertices, performing b tagging, discriminating between electrons and photons, and are the main component of the PF algorithm. Tracks have five degrees of freedom: three are needed to define the momentum vector (p_T, η, φ), and two are for the impact parameter (d_{xy}, d_z), where d_{xy} and d_z are respectively the r an z coordinate of the point of closest approach of the track of the z axis. The pattern recognition algorithm reconstructs the tracks from the hits measured by the silicon detectors (tracking). The measurement of the parameters of the tracks are described in Ref. [41] and discussed in the next sections.

The hits

The first step of tracking is the identification of the hits, i.e. the measurements of the charged particle trajectory done by the silicon sensors. The hits are measured starting from the charge collected in the sensors. Only the sensors with a significant amount of charge above the noise threshold are actually read-out (zero-suppression).

In the pixel detector, each hit is localized by a pixel cluster, that is a set of adjacent pixels with a signal, including both side-by-side and corner-by-corner pixels. The position of each hit is found with a fit to the distribution of the charge deposited within the pixel cluster. The fit is performed using a set of templates, obtained simulating the signal released by particles having different incident points and angles. The simulation takes into account also the bias on the hit position due to the asymmetric response as a

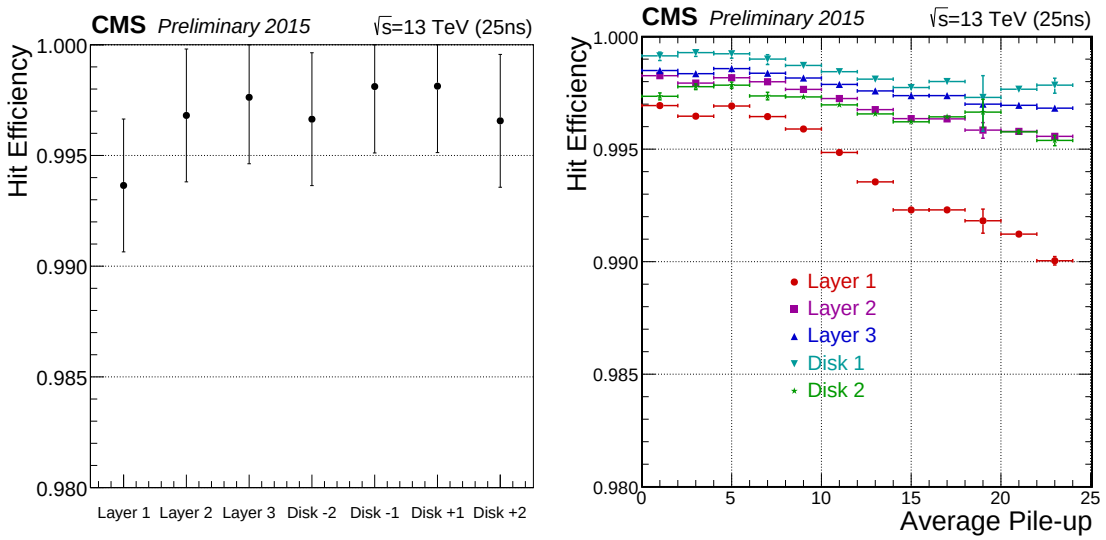


FIGURE 2.13: Pixel hit efficiency. On the left, the average pixel hit efficiency for the different pixel modules; on the right, the pixel hit efficiency as a function of the average number of pile-up collisions [47].

function of the pixel sensor depth induced by aging and radiation effects. As shown in Figure 2.13, the reconstruction efficiency of the pixel hits during Run 2 is above 99% for all layers in any condition. The small inefficiency is mainly due to failures of the read-out in very high occupancy condition.

In the strip detector, likewise to the pixel clusters, each cluster is defined as a set of adjacent strips with a signal. In order to remove potential fake hits, a zero suppression is applied both to the single strip, requiring that the collected charge is at least three times larger than the noise level, and to the whole strip cluster, requiring that the charge is larger than five times the noise. Nearby strips are added, if their strip charge is more than twice the strip noise. Eventually, each hit is determined as the charge-weighted average position of the strip cluster and is corrected for the bias due to the detector geometry (e.g. Lorentz angle).

The iterative tracking

Tracks are reconstructed using the combinatorial track finder algorithm [41]. The idea is to reconstruct tracks iteratively, starting from the tracks that are the simplest to be reconstructed (e.g. tracks with high p_T or small impact parameter). At each iteration the hits associated with the reconstructed tracks are removed from the hit collection, reducing the combinatorial of the subsequent iterations.

In details, each iteration is composed of four steps.

1. Track seeds are identified from triplets of 3D-hits or pairs of 3D-hits plus the beam spot. Only hits measured with the pixel and double strips detectors are used at

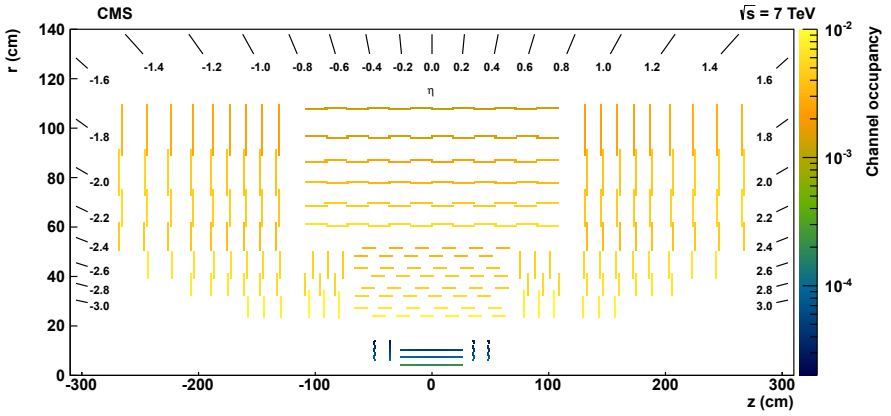


FIGURE 2.14: Tracker occupancy in events with, on average, eight simultaneous collisions [41].

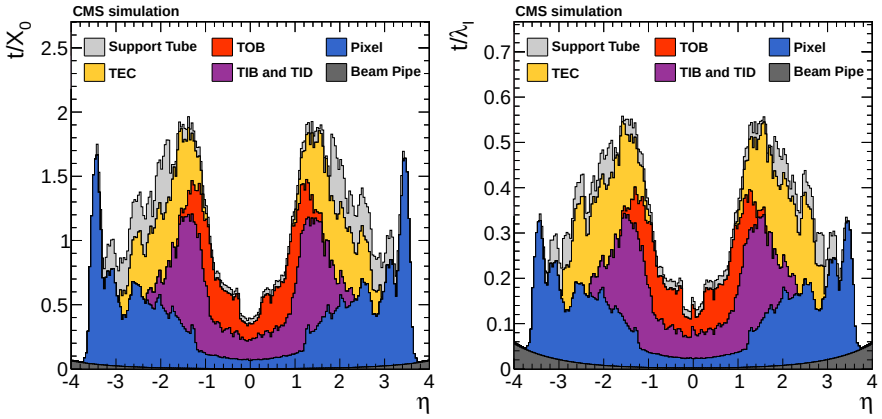


FIGURE 2.15: Total thickness t of the tracker material, as a function of pseudorapidity η, expressed in units of radiation length X_0 (left) and nuclear interaction length λ (right). The contribution to the total material budget of each of tracker subsystems is shown [41].

this stage. In particular, the pixel hits are preferred because they have higher resolution, lower occupancy and a smaller amount of material budget between the beam spot and the detector, as shown in Fig. 2.14 and 2.15. However, double strip hits are useful to recover the efficiency of displaced tracks, i.e. originating outside the beam spot.

2. The seeds are extrapolated up to the whole tracker using a combinatorial Kalman filter [48]: the hits from the different tracker layers are added to the matching tracks, and each time the track parameters are updated. In case a track is compatible with multiple hits, it is split into multiple candidates that include the different hits. In order to reduce the computational time, only the five best fitted track candidates with $\chi^2/ndof < 30$ are considered for each seed. To avoid duplicated tracks, the hits shared by multiple tracks are assigned to the track with the highest number of associated hits or, in a case of a tie, to the track with the lowest χ^2.

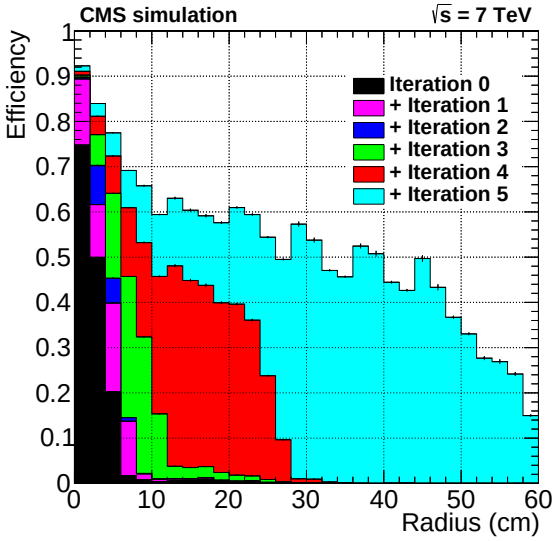


FIGURE 2.16: Cumulative contributions to the tracking efficiency from the six iterations in track reconstruction in simulated $t\bar{t}$ events, as a function of transverse distance (r) from the beam axis of the production point of each particle. Only tracks are generated with $p_T > 0.9\,\text{GeV}$ and $|\eta| < 2.5$, transverse (longitudinal) impact parameter $< 60(30)\,\text{cm}$ [41] are considered when measuring the efficiency.

3. The tracks are fitted with the associated hits considering also the effects neglected by the pattern recognition, like the non-uniformity of the magnetic field, the dependence of the hit resolution from the track parameters, and the presence of outlier hits (e.g. δ-rays).

4. Only the tracks fulfilling a minimum quality requirement are kept, the others are discarded. Eventually, the hits associated with the retained tracks are removed from the hit collection.

The track seeds used at each iteration are the following:

(Iter.0) tracks with 3 pixel hits and $p_T > 0.8\,\text{GeV}$;

(Iter.1) tracks with 2 pixel hits, close to the beam spot, and $p_T > 0.6\,\text{GeV}$;

(Iter.2) other tracks with 3 pixel hits, close to the beam spot, and $p_T > 0.075\,\text{GeV}$;

(Iter.3-5) a mixture of pixel and strip hits, to reconstruct tracks originating outside the beam spot. Usually those tracks come from photon conversions, nuclear interactions, and decays of hadrons containing s, c, or b quarks. As shown in Fig. 2.16, iteration 3-5 are crucial to recover the tracking efficiency of tracks originated with a transverse distance of $r \gtrsim 6\,\text{cm}$.

The tracking efficiency at CMS is very high. Figure 2.17 and 2.18 show the tracking efficiency for muons and pions, respectively. For muons, the efficiency is above 99% for

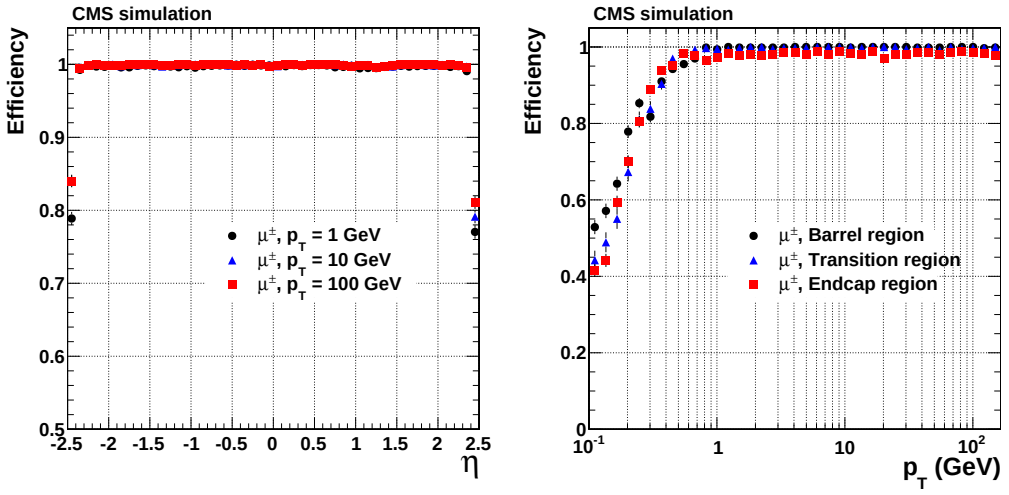


FIGURE 2.17: Tracking efficiency for muons as a function of η (left) and p_T (right) [41].

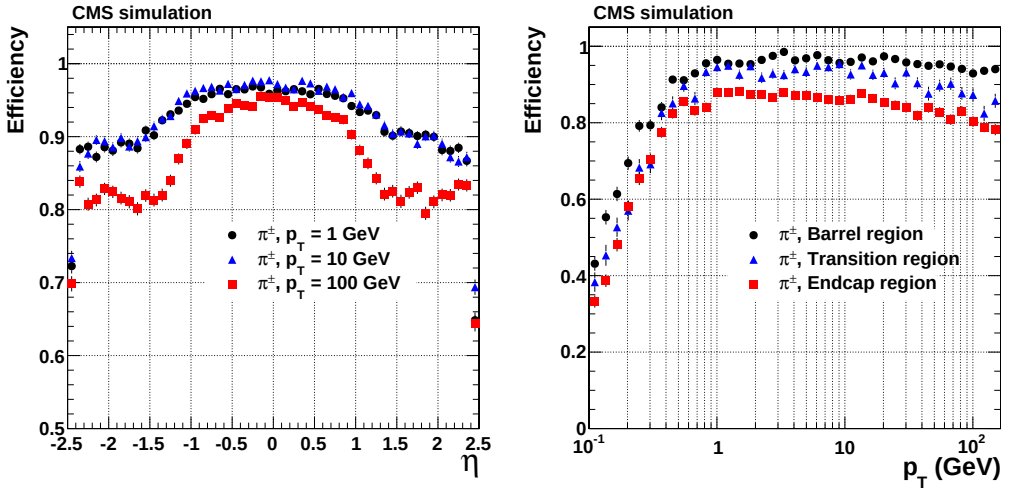


FIGURE 2.18: Tracking efficiency for charged pions as a function of η (left) and p_T (right) [41].

$|\eta| < 2.4$. For pions, the efficiency is lower because of the nuclear interaction of pions with the tracker material. Figure 2.19 shows the p_T resolution for a sample of muons as a function of $|\eta|$ and p_T. The tracking efficiency is well reproduced in simulation, as it is shown in Figure 2.20.

During the Long Shutdown 1, I worked on the improvement of tracking in the core of energetic jets. The novel algorithm is described in Appendix A. It increased the tracking efficiency, roughly, from 75% to 86% for tracks with $p_T > 2\,\text{GeV}$ in the core of jets with $p_T > 450\,\text{GeV}$.

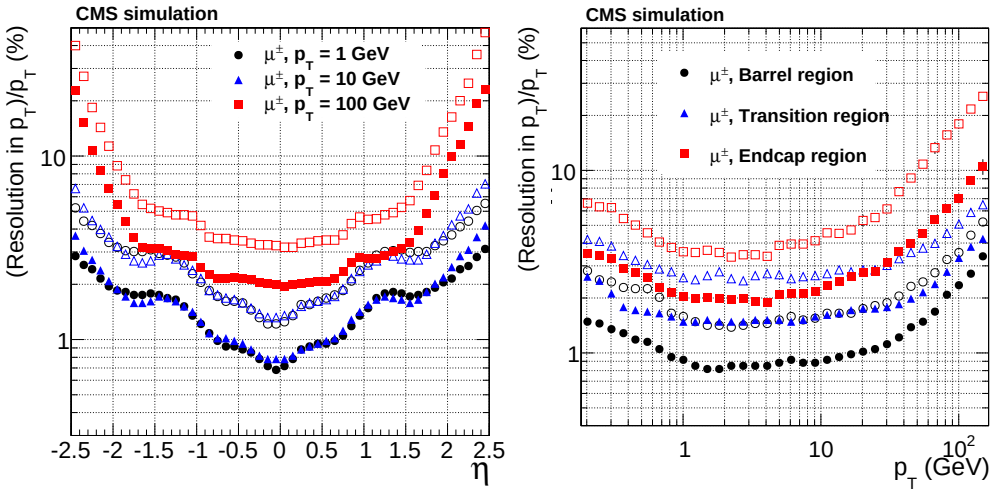


FIGURE 2.19: Tracking p_T resolution for muons as a function of η and p_T. The solid (open) symbols correspond to the resolution evaluating the error with the 68% (90%) of confidence level [41].

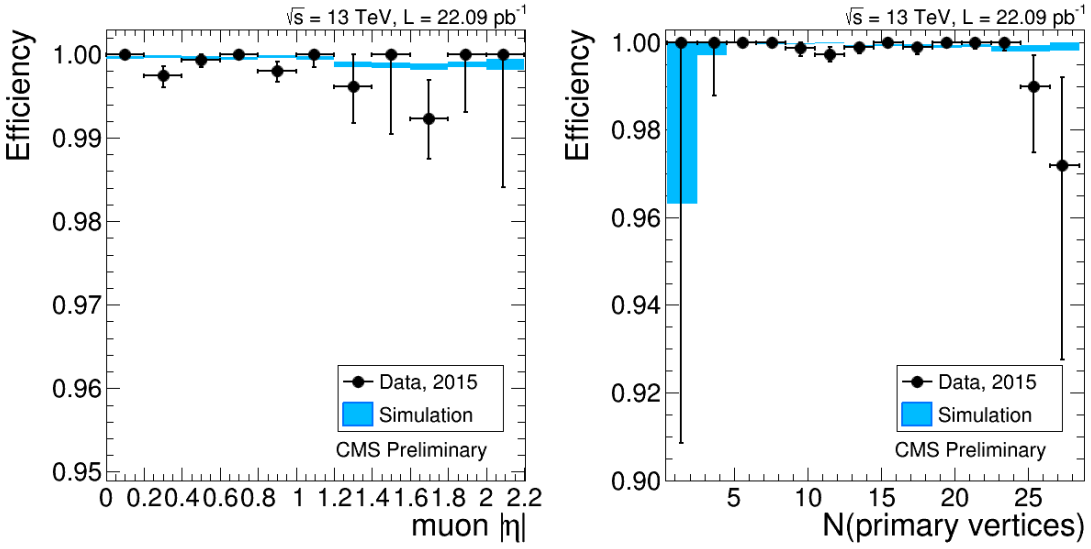


FIGURE 2.20: Tracking efficiency for muons measured with a tag-and-probe technique [49], in $Z \to \mu\mu$ events, as a function of η (left) and number of reconstructed primary vertices (right) for 2015-data (black dots) and simulation (blue bands). The tracking efficiency is well above 99% everywhere. Data shows a small efficiency drop with increasing pile-up, that is not reproduced in the simulation. It may be originated from the pile-up dependence of the pixel detector inefficiency (right plot on Fig.2.13) [50].

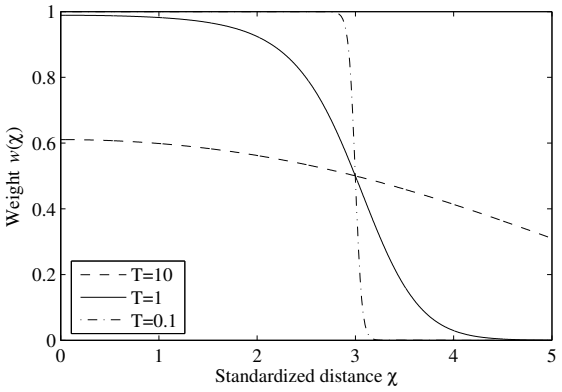


FIGURE 2.21: The function $w_i({\chi_i}^2)$ (2.4) used in the Adaptive Vertex Fitter, as a function of T and ${\chi_i}^2$, with ${\chi_c}^2 = 9$. Note that if $T = 0$ the function is a step function centered in χ_c [51].

2.3.2 The primary vertex

The primary vertices are the proton-proton collision points. As discussed in Sect. 2.1, in each proton-proton bunch crossing, there are usually tens of simultaneous proton-proton collisions (pile-up). The main algorithm used at CMS to find the position of the primary vertices is the adaptive vertex fitter [51]. This algorithm is an iterative re-weighted Kalman filter that fits a candidate vertex starting from a collection of tracks. At each iteration, the candidate vertex position and the track weights w_i are updated in two steps.

1. The candidate vertex position is determined with a fit, in the hypothesis that all tracks are originating from the same candidate vertex. In the fit, the contribution of each track is weighted by the factor w_i (at the first iteration $w_i = 1$ for every track).

2. The weights w_i are evaluated for all tracks, depending on their ${\chi_i}^2$ in the fit, using the new candidate vertex position:

$$w_i({\chi_i}^2) = \frac{1}{1 + e^{\frac{{\chi_i}^2 - {\chi_c}^2}{2T}}}, \tag{2.4}$$

   where ${\chi_c}^2$ is a constant, and T is a positive parameter. The shape of the function $w_i({\chi_i}^2)$ for some representative T values is shown in Figure 2.21.

The adaptive vertex fitter reconstructs only one vertex. In order to find the additional vertices, the algorithm is repeated using the remaining tracks that have not been associated previously with a vertex.

2.3.3 The jets

A jet is the collimated set of particles generated by the hadronization of quarks or gluons. Jets can be clustered according to several algorithms. In Run 2, the standard algorithm used by CMS is the anti-k_T with $R = 0.4$ [52]. Three kinds of jets are reconstructed at CMS: calorimetric jets, track jets, and PF jets [53].

The calorimetric jets are reconstructed using only the so-called calorimetric towers. Each calorimetric tower is made up of one or more HCAL cells and the corresponding ECAL cells. For instance, in $|\eta| < 1.4$ each tower is composed of one HCAL cell ($\Delta\eta \times \Delta\varphi = 0.0174 \times 0.0174$) and 5×5 ECAL cells. A zero-suppression is applied to clean the calorimetric towers from spurious signals. The energy resolution of the calorimetric jets is largely worse than the PF jets, that will be illustrated in Sect. 2.3.7 dedicated to the PF algorithm. However, they are still used at trigger level because they can be reconstructed very quickly since they do not need tracking.

In contrast, the track jets are reconstructed using only tracks. Tracks are required to be originating from the primary vertex. In this way, the track jet performance is independent from pile-up. On the other hand, as the neutral jet energy component is missing, the energy resolution of the track jets is worse than the calorimetric jets. Track jet are typically used only in analyses that need to be absolutely pile-up independent.

The PF jets have the best resolution and are the standard jets used at CMS. In order to reduce the pile-up contribution, PF jets are usually evaluated removing all charged particles that are originating from the pile-up vertices.

The jet energy response is corrected as a function of p_T and η applying corrections at three levels:

1. **L1 - pile up corrections** subtract the average energy deposited within the jet cone by pile-up collisions.
2. **L2 - relative jet corrections** make the jet energy response flat as a function of η. They are determined using di-jet events, assuming that these events are balanced in the transverse plane.
3. **L3 - absolute jet corrections** depend on p_T and aim to equalize the mean jet energy value equal to the energy of the parton whose hadronization produces the jet. They are calibrated in data using the balancing of Z/γ+jet events.

The corrections obtained from data and simulation are very similar. For this reason, the corrections obtained from simulation are used both in the data and in the simulation. In order to take into account the small differences between the data and simulation, residual corrections of a few percents are applied to data.

2.3.4 The missing transverse energy

The missing transverse energy (MET) exploits the conservation of the momentum in the plane perpendicular to the beam direction in order to identify the presence of particles that escape the detector without being measured. At CMS, there are three algorithms used to reconstruct the MET [54]: calorimetric MET, track MET, PF MET.

The calorimetric MET is evaluated using only the calorimeters. It is calculated as:

$$\overrightarrow{\text{MET}} = -\sum_n \left(\frac{E_n}{\cosh \eta_n} \cos \varphi_n \hat{i} + \frac{E_n}{\cosh \eta_n} \sin \varphi_n \hat{j} \right) \tag{2.5}$$

where E_n is the energy deposited in the n-th cell of the calorimeter, $\hat{i}$ and $\hat{j}$ are the versors of the x and y axis, and φ_n and η_n are the angular variables of the cell. If an event contains muons, the muon momentum p_T vector is added to the MET with negative sign and the estimated energy deposited by the muon in the calorimeters is removed from the MET computation.

The track MET, likewise the track jets, is the MET evaluated using only the tracks. It has a poor resolution since it does not include neutrals, but it is pileup independent.

The PF MET will be described in the section about the PF algorithm (Sect. 2.3.7).

One of the main corrections applied to the MET is the propagation of the L2 and L3 jet energy corrections described in Sect. 2.3.3. The idea is to separate the MET originating from jets with $p_T > 10\,\text{GeV}$ and improve the MET resolution reevaluating the MET after applying the L2 and L3 corrections on those jets. Note that the L1 corrections are not applied. The reason is that the pile-up collisions are supposed to be isotropic and without MET. On the contrary, the angular distribution of the jets in an event is typically anisotropic and so the application of the L1 corrections to the jets would generate fake MET. Differently from MET, the missing transverse energy evaluated on jet ($\overrightarrow{\text{MHT}} = -\sum_i \text{jet}\ \vec{p}_T^{\,i}$) is corrected using the L1 corrections too.

2.3.5 The isolated photons and electrons

Both electrons and photons [55, 56] release essentially all their energy in the electromagnetic calorimeter. As electrons are charged, their signature includes also a high momentum track pointing towards the energy deposit. On the contrary, photons are neutral and hence they cannot be detected with the tracker, except the case of the photon conversion $\gamma \to e^+e^-$ in the tracker material. The reconstruction of isolated electrons and photons is seeded by the identification of a cluster of 5×5 ECAL crystals with a large energy deposit. The cluster is then extended, along the φ coordinate, into

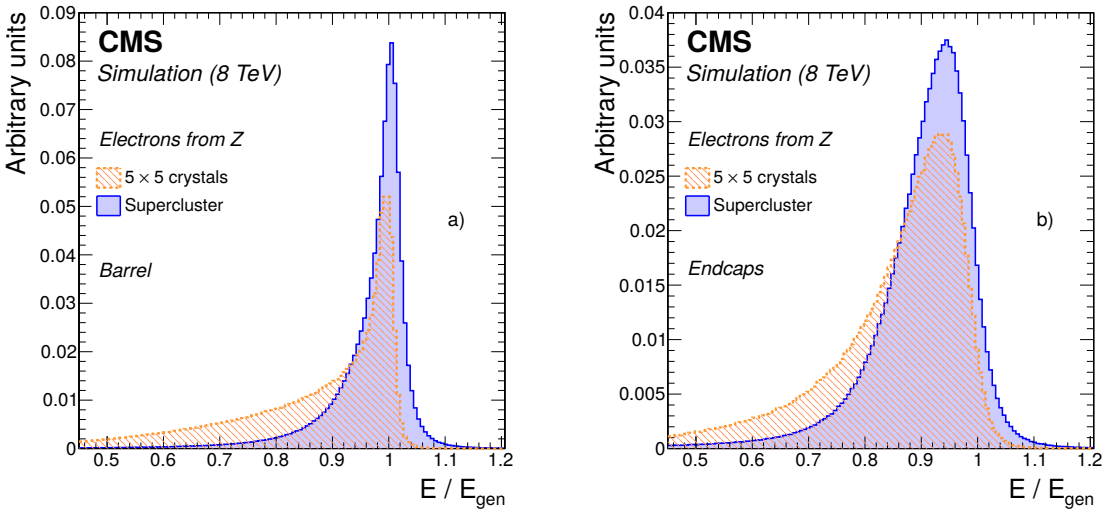


FIGURE 2.22: Ratio of the energy deposited in an ECAL cluster (yellow) and in a super-cluster (blue) to the true energy of the generated electron. The plots have been produced both for the barrel (left) and end-cap (right). No energy correction is applied to any of the distributions [55].

a super-cluster in order to collect also the energy emitted by radiating electrons or converted photons. Figure 2.22 shows the improvement in the electron energy resolution obtained using the super-cluster instead of the standard 5×5 cluster.

The tracking of the converted photons and primary electrons is performed with a dedicated algorithm called Gaussian sum filter (GSF) [57]. This algorithm is needed to recover the tracking efficiency when strong bremsstrahlung radiation is present. On the other hand, the GFS tracking is very slow and hence is run only around the isolated photon/electron candidates, identified as ECAL clusters with an HCAL/ECAL energy deposit ratio in $\Delta R < 0.15$ less than 15%. This tracking algorithm allows also to measure the energy loss due to bremsstrahlung radiation by comparing the p_T of the track measured near the primary vertex and near the calorimeter. Figure 2.23 shows the distribution of the number of hits of an electron track reconstructed with the GSF and the standard Kalman Filter tracking. As expected, electron tracks reconstructed with the GSF algorithm have more associated hits because this algorithm is more efficient in the presence of strong bremsstrahlung radiation.

Tracking is also useful to improve the energy resolution of soft electrons. Figure 2.24 is a comparison of the electron energy resolution measured combining the GSF track p_T with the super-cluster energy (E_{SC}) and using the E_{SC} only. The tracking improves significantly the energy resolution for electrons with $p_T \lesssim 20\,\text{GeV}$.

An important issue that affects both photons and electrons is the change of the ECAL crystal transparency during the data taking induced by the radiation. This effect is measured and corrected using a system of optical fibers, which bring to the crystals a standard laser pulse that measures the ECAL transparency in-situ. Figure 2.25 shows the ratio between the electron energy measured in the calorimeter and the momentum

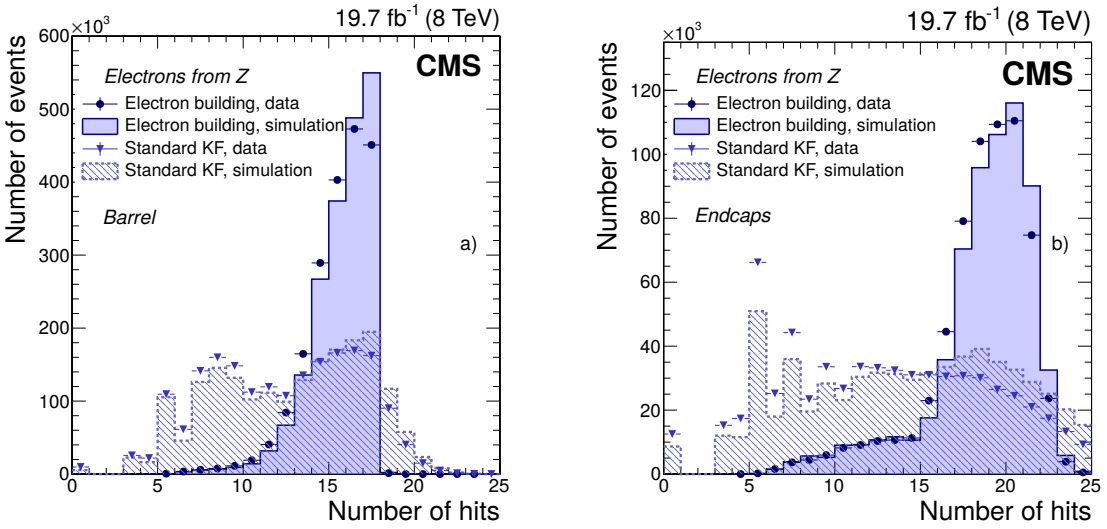


FIGURE 2.23: Number of hits per electron track reconstructed with the GFS (solid line, dark blue dots) and standard Kalman filter tracking (dashed line, light blue dots). The plot has been obtained both for data (dots) and simulation (line) and both for the barrel (left) and end-cap (right). As expected, the GFS tracking recovers a large part of the hits missed by the standard reconstruction of the electron tracks [55].

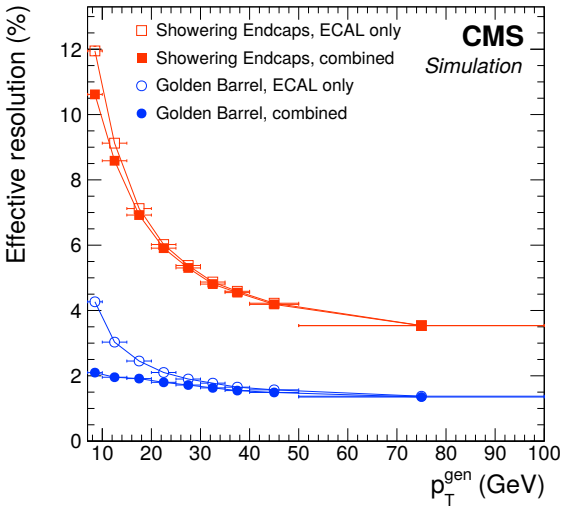


FIGURE 2.24: Electron momentum resolution after combining the super-cluster energy (E_{SC}) with the tracker p_T (solid symbols) and using only E_{SC} (open symbols). The plots have been produced for two categories: electrons in the barrel with little bremsstrahlung radiation (circles) and electrons in endcap with high bremsstrahlung radiation (squares) [55].

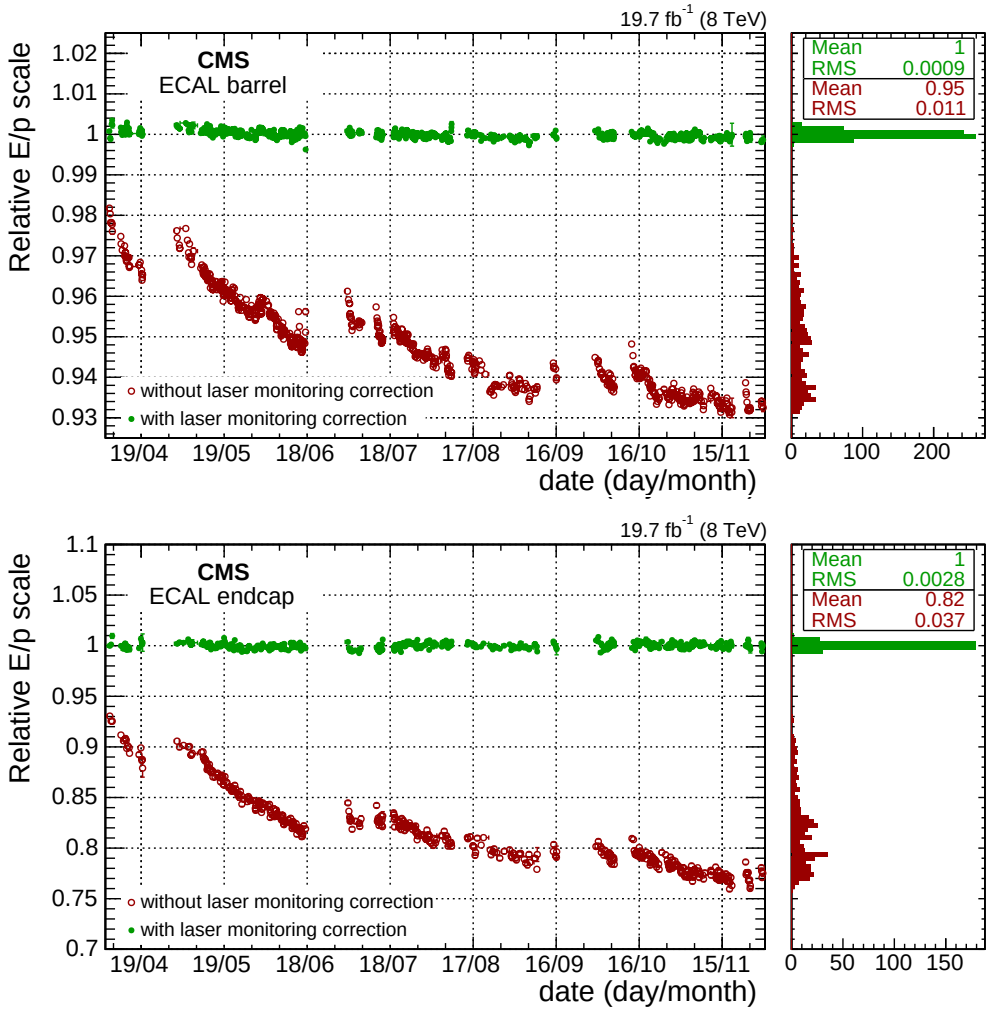


FIGURE 2.25: Ratio of the electron super-cluster energy to the electron track p_T, as a function of the date at which the events have been recorded. The plot have been produced before (red points) and after (green points) the laser monitoring corrections have been applied and both for the barrel (top) and end-cap (bottom). Note the different scale of the two plots [56].

measured in the tracker during the data-taking before and after the laser monitoring corrections have been applied. Figure shows that the E/p, after the application of the laser correction, is kept constant and close to 1 during the whole data-taking.

The main variables used to discriminate electrons and photons from the backgrounds are the ratio between the ECAL and HCAL energy deposits, that reduces the hadronic background, and the transverse shape of the shower along the η coordinate, that is useful to reject the $\pi^0 \to \gamma\gamma$ background. In the case of electrons, additional discriminating variables are the distance between the electron track and the ECAL supercluster barycenter, and the ratio between the track momentum and the energy deposited in the supercluster. Another important background for electrons are the converted photons. It is reduced vetoing the presence of tracks compatible with the photon conversion

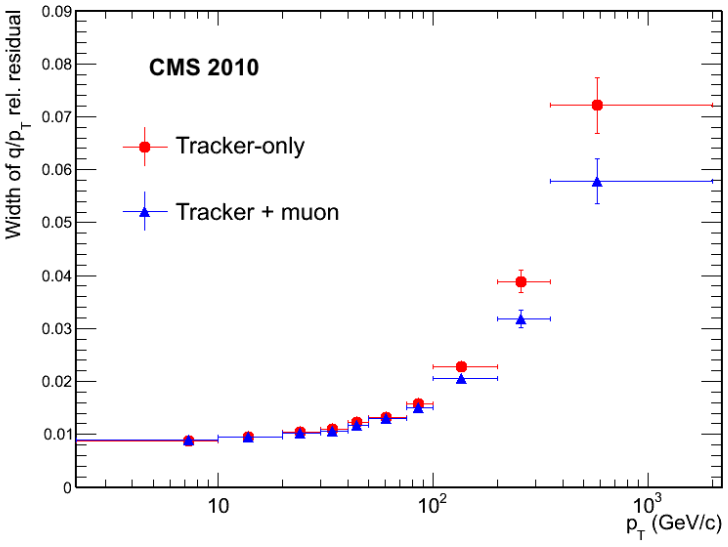


FIGURE 2.26: Widths of Gaussian fits to the distributions of the muon q/p_T relative residuals, as a function of the momentum, for tracks obtained using the tracker only (red) or the tracker plus the muon chambers (blue) [58].

$\gamma \to e^+e^-$, and requiring that the electron track contains a hit in the innermost available tracker layer, in order to restrict the region where the photon may have generated the non-prompt electron.

In addition to the isolated photons and electrons, the reconstruction of non-isolated electrons and photons is important for many purposes like the improvement of the jet energy resolution and b tagging. It will be described in the section dedicated the PF algorithm.

2.3.6 The muons

Muons are the only charged particles that can escape from the CMS detector. Muons signature is characterized by measurements in the muon chambers, outside the superconducting solenoid. In the standard muon reconstruction, the muon tracks are first reconstructed independently in the inner tracker (tracker track) and in the muon chamber (standalone-muon track). Then, two reconstruction approaches are used:

- **Global muon reconstruction** (outside-in). Each standalone-muon track is matched with a tracker track. The global-muon track is then fitted with a Kalman filter combining the hits of the tracker and standalone-muon tracks. Figure 2.26 shows that the global-muon fit improves the momentum resolution starting from $p_T \gtrsim 200\,\text{GeV}$ with respect to the tracker-only fit.

- **Tracker muon reconstruction** (inside-out). In this approach, all tracker tracks with $p_T > 0.5\,\text{GeV}$ and total momentum $p > 2.5\,\text{GeV}$ are considered as possible muon candidates and are extrapolated to the muon chambers. If at least one matching muon segment is found, the tracker track is tagged as a tracker muon. The extrapolated track and the segment are considered matching if the distance between them in the muon chamber coordinate is less than 3 cm, or if the pull $|x_{meas} - x_{extrap}|/\sigma(x)$ is less than 4, where $\sigma(x)$ is the uncertainty on the extrapolated track position in the muon chamber x_{extrap}.

There are different muon quality selections according to the efficiency/purity desired.

- **Soft muon** selection. It is the loosest selection: tracker-muons are required to have a pull less than 3 between the extrapolate tracker-track and the muon chamber hit.
- **Tight muon** selection. The global-muon is required to have ≥ 2 hits in the muon stations, ≥ 10 hits in the tracker (including ≥ 1 pixel hit), a transverse impact parameter less than 2 mm, and $\chi^2/ndof < 10$.
- **PF muon** selection. In the PF algorithm, a specific muon selection is used. It is described in the following section about the PF algorithm.

2.3.7 The particle flow

The particle flow (PF) is an algorithm that combines the information of all the CMS sub-detectors in the attempt to reconstruct and identify the particles produced in the proton-proton collision, categorizing them in charged hadrons, neutral hadrons, photons, electrons, and muons. These five categories of particles are then used to reconstruct more complex objects, like jets and MET. The PF is able to enhance significantly the reconstruction performance. In the PF, the p_T of soft charged particles is measured using the tracker, instead of the calorimeters, increasing largely the resolution. The global improvement in the jet reconstruction is substantial both in terms of the fraction of the energy measured (response) and resolution, as it is shown in Fig. 2.27. The improvement in the MET resolution is reported in Fig. 2.28.

The first step of the PF algorithm is the iterative tracking (Sec. 2.3.1) and the reconstruction of the HCAL and ECAL topological clusters, that are groups of HCAL and ECAL cells with a significant energy deposit. Afterwards, the tracks are extrapolated from the tracker to the calorimeters and to the muon chambers. Here, the tracker and muon tracks and the topological clusters matching in the (η, φ) plane form a PF block. Figure 2.29 shows an example of PF block for a simulated jet.

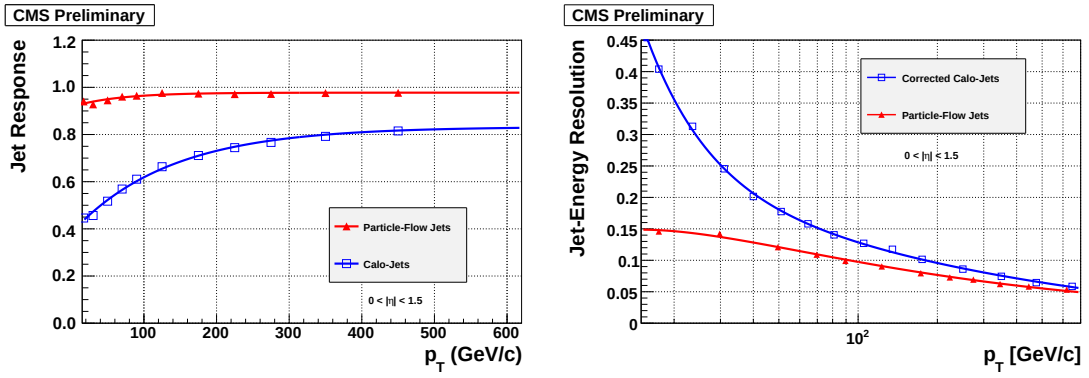


FIGURE 2.27: Comparison between the reconstruction of calorimetric and PF jet using simulation. In particular, on the left, the jet response, and on the right, the jet-energy resolution as a function of the jet p_T [46].

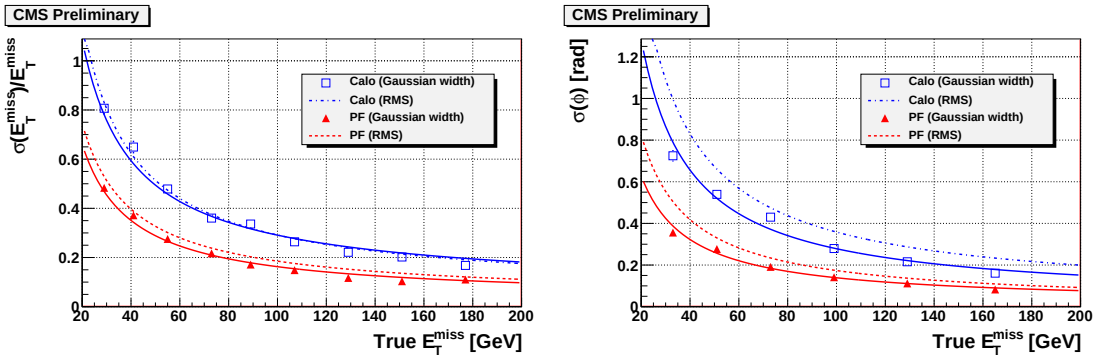


FIGURE 2.28: MET energy (left) and angular (right) resolution as a function of the generated MET. The performance is shown both for the calorimetric (blue) and PF (red) MET. The MET resolutions have been measured as Gaussian widths (solid line and markers) and root mean square (dashed line). The plots have been obtained using a sample of $t\bar{t}$ simulated events [46].

The real reconstruction and identification of particles take place within each PF block. At the beginning, muons are reconstructed according to the algorithm described in Sec. 2.3.6 and all tight muons are directly tagged as PF muons. Every isolated global muon is identified as PF muon as well. The isolation requirement is that the sum of the track momenta and calorimeter energy deposits within $\Delta R < 0.3$ from the candidate muon have to be lower than 10% of the muon p_T. Tracks associated with the PF muons are then removed from the PF block, and the expected energy released in the calorimeters by muons is taken into account in the subsequent reconstruction. Loose muons are reconstructed later, using the objects not assigned during the PF reconstruction.

Afterwards, electrons are reconstructed in a way similar to the standard electron reconstruction described in Sec. 2.3.5, but here it has been optimized to reconstruct also non-isolated electrons that may be contained in a jet. The PF electron reconstruction is useful to improve the jet energy resolution, in particular in the case of b jets containing a B hadron decaying into a soft electron. In the PF, differently from the standard electron reconstruction, the GSF tracking - the algorithm used to reconstruct the electron

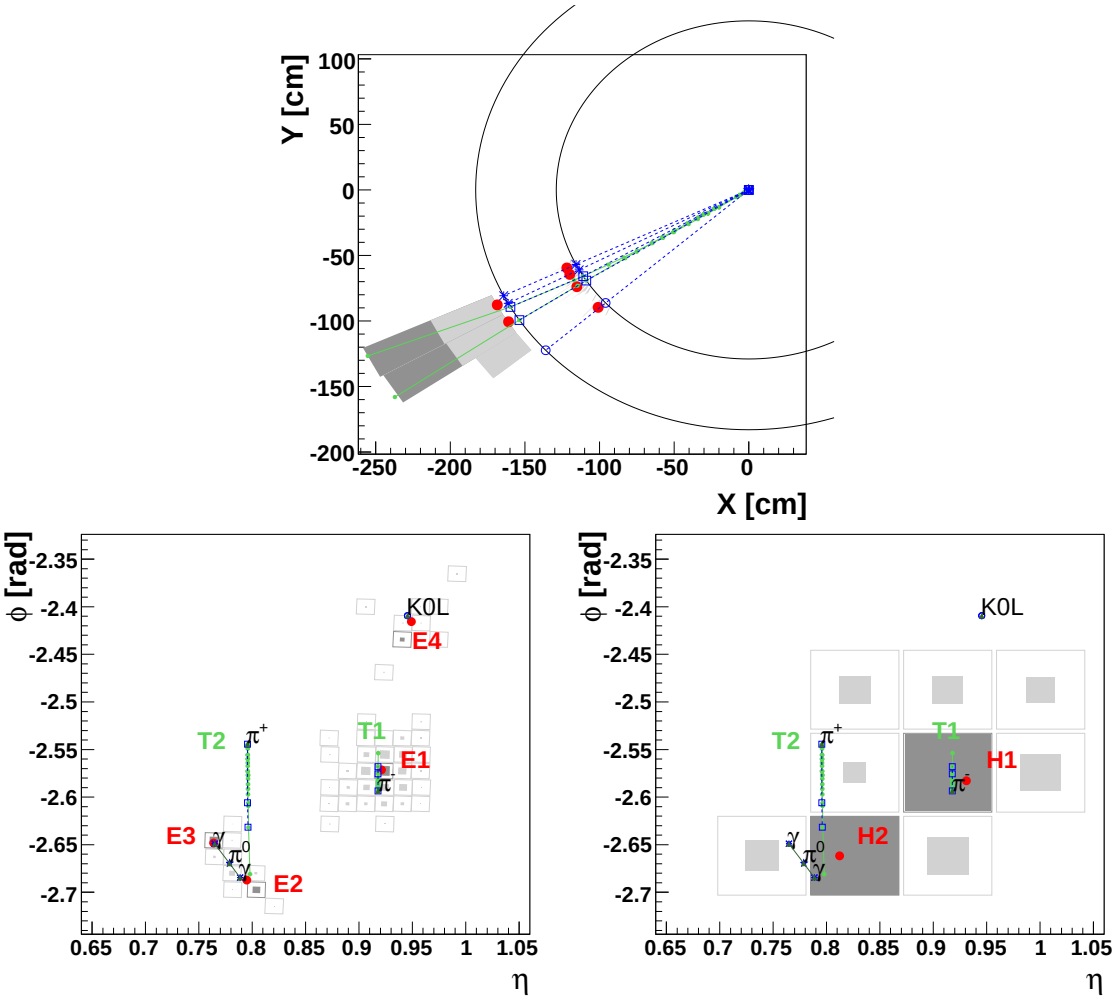


FIGURE 2.29: A PF block reconstructed in a simulated jet containing the following particles: π^0, π^+, π^-, K^{0L}. The light blue squares indicate the places where the generated particles reach the ECAL and HCAL. The green lines, T1 and T2, correspond to the tracks of the charged particles reconstructed by the tracker and extrapolated up to the calorimeters. On the top, the PF block is shown in a transversal section of CMS. On the bottom left (right), the energy deposits in the ECAL (HCAL) cells are shown along with the reconstructed ECAL (HCAL) clusters. In this PF block, four ECAL clusters (red circles E1, E2, E3, E4) and two HCAL clusters (H1, H2) were reconstructed [46].

track - is run without any requirement on the isolation. Photons tangent to an electron track are considered part of the bremsstrahlung radiation, as illustrated in Fig. 2.30, and hence their energy is assigned to the electron. Finally, a BDT is used to distinguish electrons from neutral pions. The BDT uses both the calorimetric and tracker variables, like the fraction of the ECAL and HCAL energy deposits, the shape of the ECAL cluster along the η coordinate, the ratio between the track momentum and the ECAL cluster energy, and the fraction of energy loss by bremsstrahlung.

Finally, charged and neutral hadrons, and photons are reconstructed. If multiple tracks point to the same HCAL cluster, all links are retained. In contrast, if more HCAL clusters match the same track, only the cluster closer to the track is linked. For ECAL

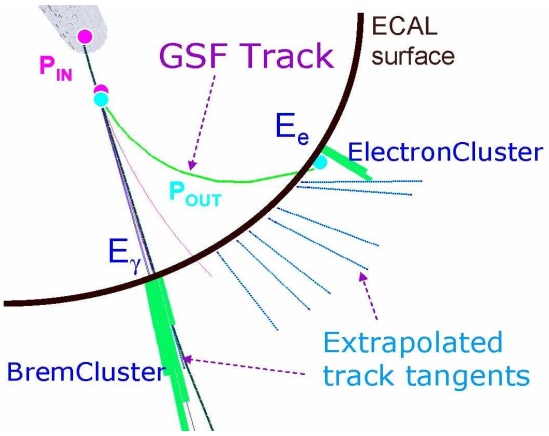


FIGURE 2.30: Sketch illustrating a bremsstrahlung photon emitted from an electron. In the PF algorithm, bremsstrahlung photons are identified as the ECAL cluster tangent to the electron track. The energy of the bremsstrahlung photons is added to the electron energy [59].

clusters, similar criteria are used, but in the case of multiple ECAL clusters pointing towards the same track they are linked to the track only if they are compatible with a shower fluctuation.

If the total energy deposited in the calorimeters clusters is lower than three standard deviations from the corresponding track momenta, loose muons are searched for and eventually removed from the PF tracks collection. If the transverse energy is still lower than the track p_T, the tracks with the larger uncertainty are removed as long as the two values are different. Instead, if the calorimetric energy is higher than one standard deviation from the track p_T, neutral objects are added (photon or neutral hadron), according to the HCAL and ECAL energy deposits. A detailed description of the PF algorithm is in Ref. [46].

2.3.8 The b tagging

Jets originated from the hadronization of b quarks contain B hadrons. The life-time of a B hadron is around 10^{-12} s. A B hadron with $p \sim 50$ GeV decays displaced from the primary vertex of ~ 3000 μm. The presence of a displaced secondary vertex is the main characteristic used to identify b jets. In case no secondary vertices are reconstructed, tracks with high impact parameters are used for b tagging. Figure 2.31 illustrates this concept.

The simplest b-tag algorithm used by CMS is the track-counting algorithm. It is based on the measurement of the impact parameters of the tracks inside the jet. Figure 2.32 shows how the impact parameter is defined. Approximating the B hadron direction with the jet direction, the secondary vertex corresponds to the point on the jet axis with the minimum jet-track distance, that is represented in Fig. 2.32 by the point Q. The point S,

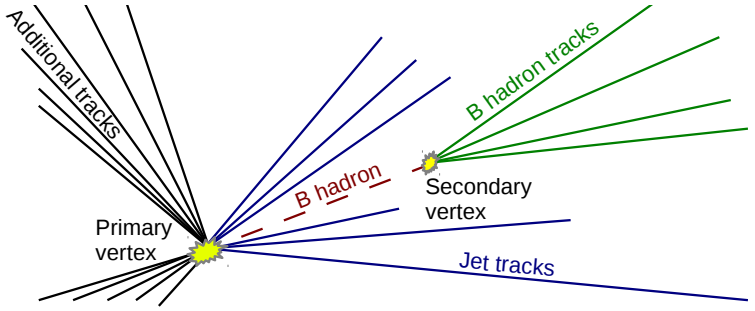


FIGURE 2.31: Sketch of an event containing a b jet. Among the particles originated from the proton-proton collision (primary vertex), a B hadron is produced and it decays in a displaced vertex (secondary vertex). The presence of the secondary vertex allows to tag the b jet. In case no secondary vertices are reconstructed, the b jet can be still tagged by the presences of tracks with high impact parameter.

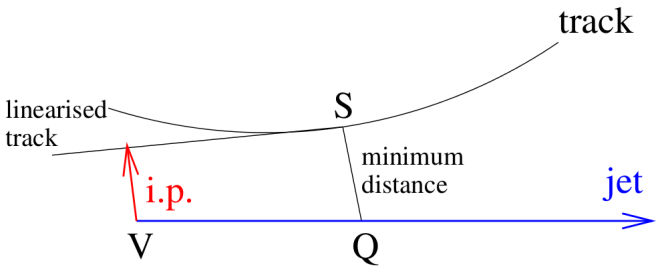


FIGURE 2.32: Definition of the impact parameter [60].

instead, is the point, along the track, with the minimum distance from the jet axis. The minimum distance between the point V and the tangent to the track, passing through the point S, defines the impact parameter. Usually the impact parameter is evaluated 3D, but it can be evaluated also using only the transverse plane. In this case, it is called 2D impact parameter. The sign of the impact parameter is defined as follows: if the projection of the impact parameter along the jet direction has the same direction of the jet, then the impact parameter is considered positive, otherwise it is negative. The sign is useful to improve the discriminating power of the impact parameter. Indeed, tracks originating from the B hadron decay are expected to have a direction close to the jet direction, and therefore to have a positive sign. The impact parameter is independent, at first order, from the momentum of the B hadron because the angle between the track and jet directions is roughly proportional to $1/\gamma$, and the displacement of the secondary vertex is proportional to γ.

A good variable to tag b-jets is the significance of the impact parameter:

$$S = \frac{IP}{\sigma_{IP}} \tag{2.6}$$

where IP is the impact parameter, and σ_{IP} is its uncertainty. Figure 2.33 shows the distribution of the signed 3D impact parameter significance in 13 TeV data compared with

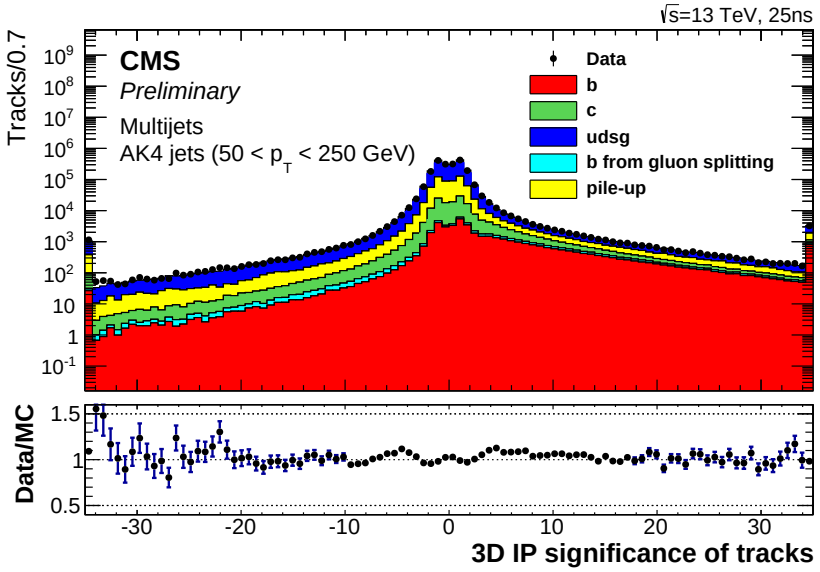


FIGURE 2.33: The 3D signed impact-parameter significance for all selected tracks associated with jets of 50 GeV $< p_T <$ 250 GeV in 13 TeV data. The small observed disagreement between data and MC is attributed to different alignment scenarios [61].

the expectation from simulations. As expected, tracks with higher impact-parameter significance have a higher probability to be originated from b jets.

The track counting is a b-tagging algorithm that tags jets depending on the n-th highest impact-parameter track significance. There are a few track-counting algorithms: the track-counting very-high efficiency ($n = 1$, TCVHE), high efficiency ($n = 2$, TCHE), and high purity ($n = 3$, TCHP). Figure 2.34 shows the track-counting high-purity discriminant distribution in multi-jet and $t\bar{t}$ enriched samples. A natural extension of the track-counting algorithms is jet probability (JP) that combines the probabilities of each track to originate from the primary vertex. If the track probability is lower than 0.5%, it is set to 0.5%. On average, the B hadron decays in four reconstructed tracks. An improved version of the JP takes into account only the four tracks with the largest impact parameter and it is called jet B probability (JBP). Figure 2.35 shows the jet-probability discriminant distribution in multi-jet and $t\bar{t}$ enriched samples.

Another b-tagging method is the search for displaced secondary vertices originated from B hadron decays. After the identification of the primary vertex, tracks not compatible with the vertex and close to the jet direction are selected. Then, the Adaptive Filter Vertex (Sect. 2.3.2) is run again using the selected tracks in the attempt to reconstruct the secondary vertex. If the new vertex fulfills certain quality requirements, it is retained and used for b tagging. Similarly to the track-counting, the significance of the distance between the secondary and primary vertex can be used as a b-tag discriminant. Depending on the number of tracks associated with the secondary vertex, this algorithm is called simple-secondary vertex high-efficiency ($n = 2$, SSVHE) or high-purity ($n = 3$, SSVHP).

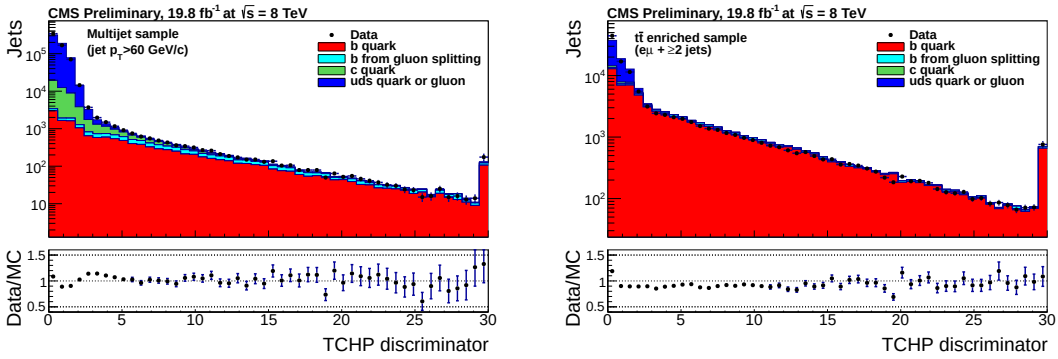


FIGURE 2.34: Track counting high purity discriminant in multi-jet and $t\bar{t}$ enriched samples. The discriminant is defined as the third highest impact-parameter significance evaluated on tracks associated with the jet [62].

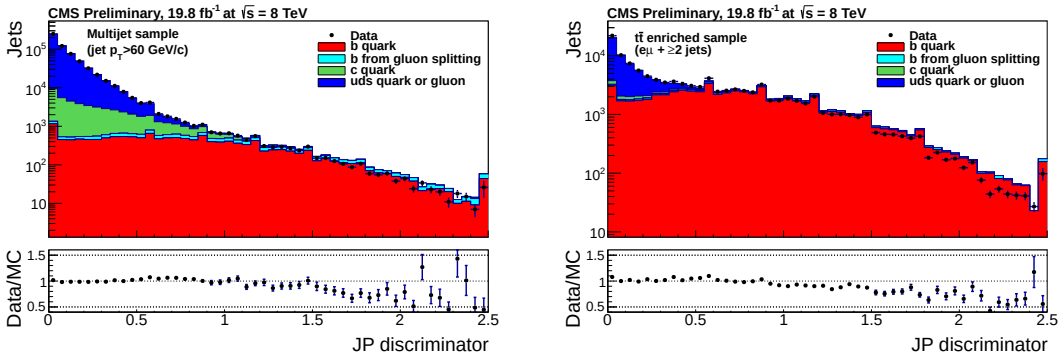


FIGURE 2.35: Jet probability discriminant in multi-jet and $t\bar{t}$ enriched samples. The discriminant is a combination of the probabilities of each track to originate from the primary vertex. The bumps in the distributions are due to the minimum value of the track probability, that is set to 0.5% [62].

The standard b-tagging algorithm used by CMS is the combined secondary vertex (CSV) that is a multivariate discriminant that combines the following variables:

- 2D distance between primary and secondary vertex;
- invariant mass of charged tracks compatible with the secondary vertex;
- number of tracks at the vertex;
- ratio of the energy carried by tracks at the vertex with respect to all tracks in the jet;
- pseudorapidities of the tracks at the vertex with respect to the jet axis;
- significance of 2D and 3D tracks impact parameter respect to the primary vertex;
- number of tracks in the jet.

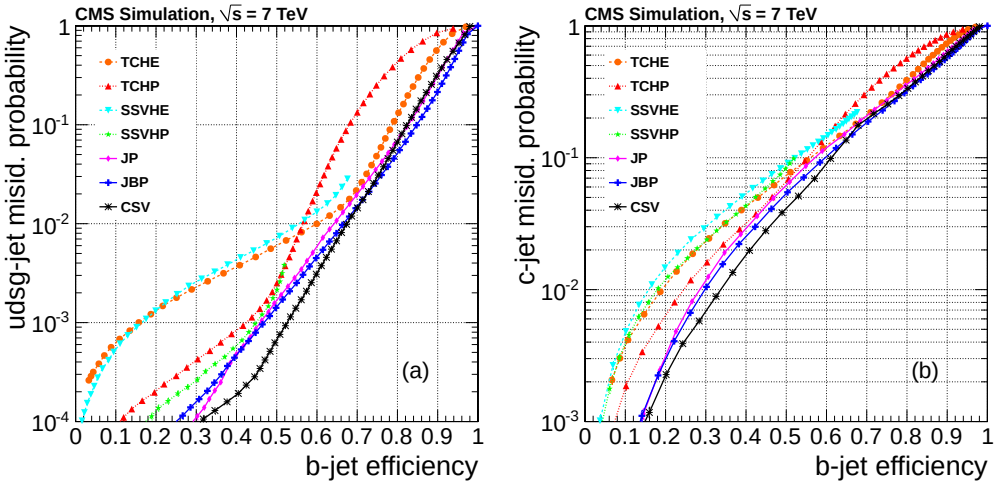


FIGURE 2.36: B-tagging performance for the different b-tagging algorithms. The plots show the misidentification probability of uds (left) and c (right) jets as a function of the b-tagging efficiency. The black line corresponds to the performance of the standard b-tagging algorithm (CSV). The plot has been produced using jets with $p_T > 60\,\text{GeV}$ of a multi-jet simulation [63].

The CSV discriminant is combined evaluating two likelihood ratios to discriminate b jets from c jets and $udsg$ jets. The two discriminants are then combined with prior weights of 0.25 and 0.75, respectively.

Figure 2.36 is a comparison of the performance of the b-tagging algorithms used at CMS.

Run-2 improvement

In 2015, the search for secondary vertices has been improved using the inclusive vertex finder (IVF). The new algorithm identifies the secondary vertices in a way completely independent from the jet reconstruction. The IVF is particularly useful to perform b tagging in topology where two or more jets overlap between them (boosted topology). In addition, during Run 2, the combined secondary vertex algorithm has been improved, using a neural network instead of the likelihood ratio. This new version is called CSVv2. Figure 2.37 shows the distribution of the CSVv2 in $t\bar{t}$-enriched and multijets regions. Figure 2.38 shows the improvement obtained by the IVF and by the neural network in the b tagging of jets with $p_T > 30\,\text{GeV}$ in a $t\bar{t}$ simulation.

Usually, in the data analysis, three b-tagging working points are defined, depending on the fake rate: loose (10%), medium (1%), and tight (0.1%). Table 2.1 reports the discriminator values to be used in the Run 2 for the three working points.

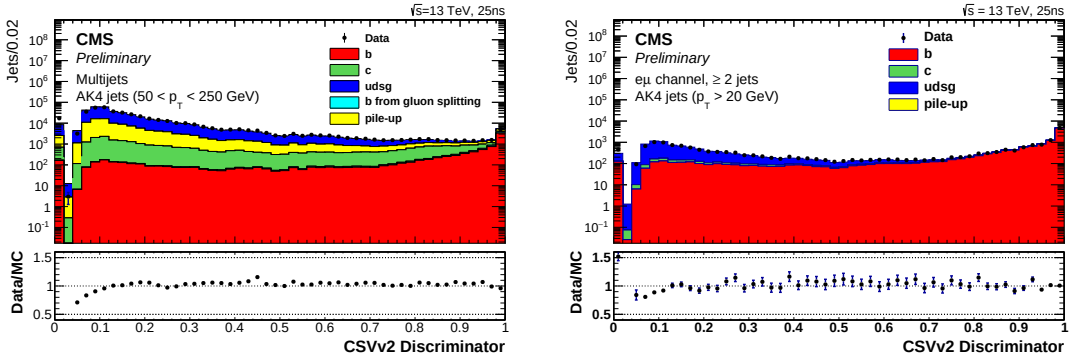


FIGURE 2.37: New combined secondary vertex discriminant used in 13 TeV data. The distribution has been obtained in multi-jet (left) and $t\bar{t}$ (right) enriched samples [61].

Scale factors

The b-tagging is not perfectly reproduced by simulations. For this reasons, scale factors are applied to simulations in order to match the data [61]. The scale factors are applied in bins of jet p_T, η, CSVv2, and depending on the jet flavor. The scale factors are obtained with a tag-and-probe technique, in $t\bar{t}$ and $Z+$jets enriched regions. Alternative methods use multijet events with a muon inside one jet as control region.

The $t\bar{t}$-enriched control region is defined requiring exactly two isolated leptons (muon or electron) with opposite sign, Z mass veto, MET, and exactly two jets. The tag jet is required to have CSVv2 > 0.800 (misidentification $< 1\%$). In this way, a large fraction of probe jets are b jets.

The $Z+$ jets-enriched control region is defined requiring exactly two isolated leptons (muon or electron) with opposite sign, same flavor and compatible with the Z mass, and exactly two jets. The tag jet is required to have CSVv2 < 0.460 (b-jet rejection $\gtrsim 80\%$). In this way, a large fraction of probe jets are light jets.

The scale factors are obtained fitting the light-, c-, and b-jets normalizations of the CSVv2 distributions of the probe jet in bins of p_T and η, in $t\bar{t}$ and $Z+$ jets control regions. An example of such distributions is shown in 2.39. The scale factors have been tested in a $t\bar{t}$ semi-leptonic control regions, defined as an isolated lepton, two jets

TABLE 2.1: Working points for CSVv2 in Run 2.

working point	discriminator value	b efficiency	fake rate
CSVv2 loose	0.460	$\approx 83\%$	$\approx 10\%$
CSVv2 medium	0.800	$\approx 69\%$	$\approx 1\%$
CSVv2 tight	0.935	$\approx 49\%$	$\approx 0.1\%$

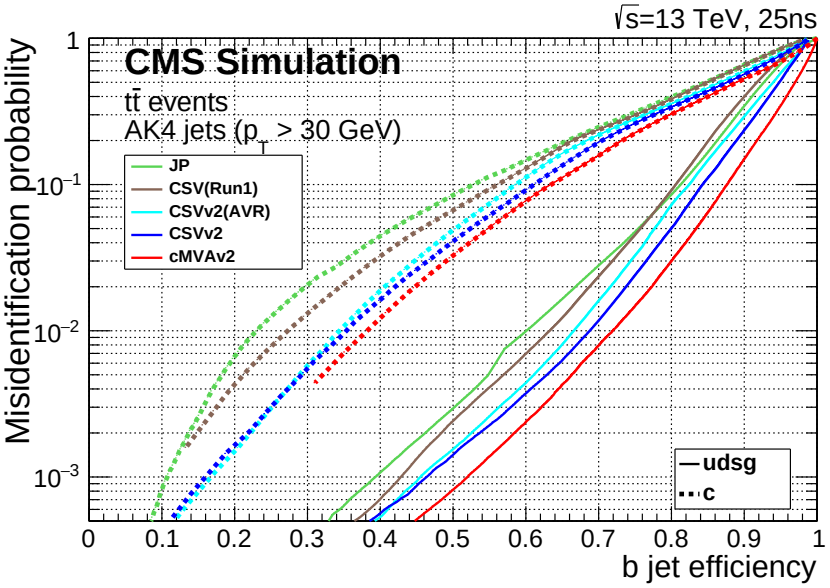
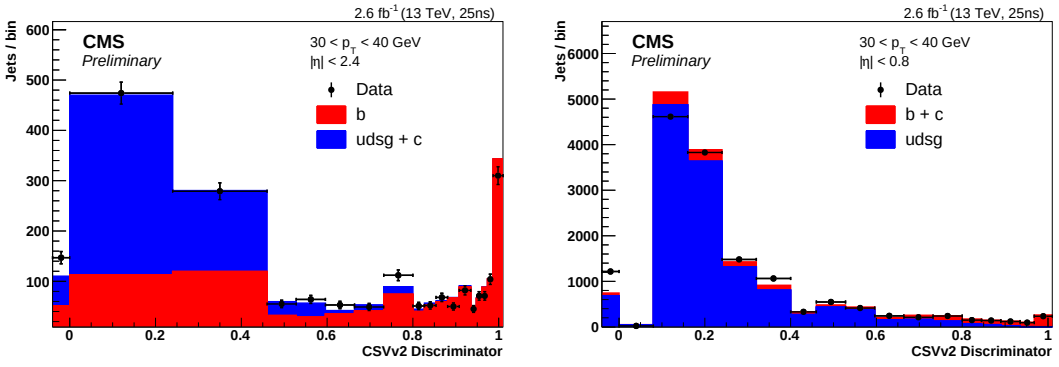


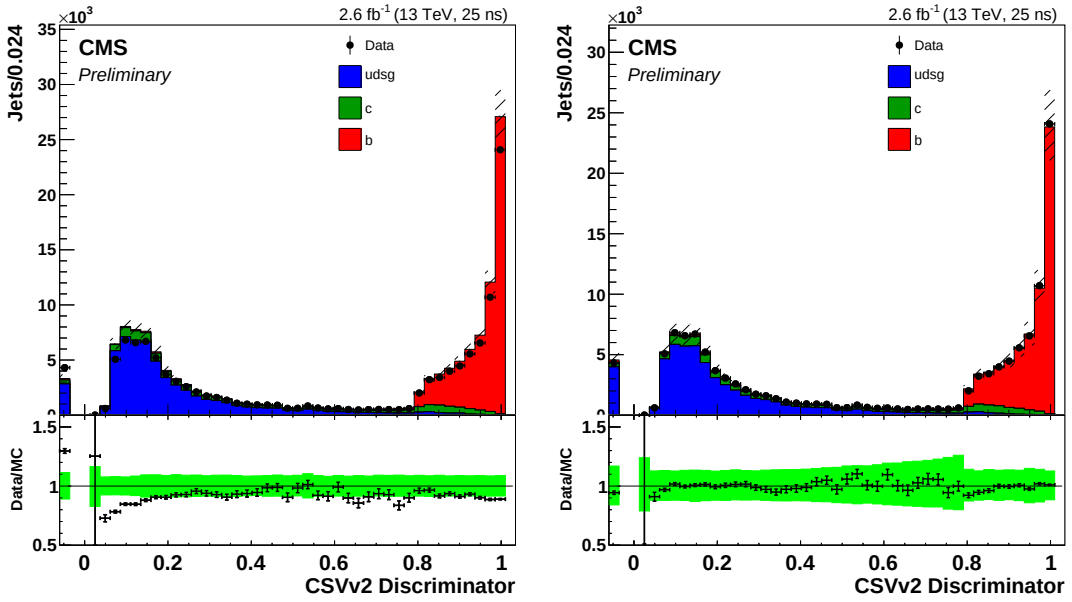
FIGURE 2.38: Misidentification probability of light jets (solid line) and c jets (dashed line) as a function of the b-tagging efficiency for jets with $p_T > 30\,\text{GeV}$ in a $t\bar{t}$ simulation. The performance are shown for the following algorithm: Jet Probability (JP), Combined Secondary Vertex based on the likelihood ratio (CSV Run 1), Combined Secondary Vertex based on the BDT (CSVv2), Combined Secondary Vertex based on the BDT and on the IVF (CSVv2), and a multivariate discriminant that exploits also the soft lepton decay (cMVAv2) [61].



(A) $t\bar{t}$ (fully leptonic) (B) Z+jets

FIGURE 2.39: Combined secondary vertex discriminant for the probe jet in $t\bar{t}$ (left) and Z+jets (right) enriched samples, before the application of the scale factor, for jets with $30\,\text{GeV} < p_T < 40\,\text{GeV}$, and $|\eta| < 2.4$ (in $t\bar{t}$) and $|\eta| < 0.8$ (in Z+jets) [61].

with CSVv2 > 0.800 and two additional jets. Figure 2.40 shows the CSVv2 distribution before and after applying the scale factors.



(A) without scale factors (B) with scale factors

FIGURE 2.40: Combined secondary vertex discriminant for jets in $t\bar{t}$ semi-leptonic decay control region, before (left) and after(right) applying the scale factors. The green band around the data-to-simulation ratio includes both statistic and systematic uncertainties, including the uncertainty on the measured scale factors [61].

2.4 Trigger

The beam revolution time of LHC is about $T_{rev} \sim 90\,\mu$s. Considering that in 2015 each proton beam was containing up to 2244 bunches, the corresponding bunch-crossing rate was around $R_{ZeroBias} = 2244/90\,\mu\text{s} \sim 25\,\text{MHz}$. In future, the bunch-crossing rate is expected to be close to the design value of $R_{ZeroBias} \sim 30\,\text{MHz}$ (2808 bunches). Saving the whole information generated by all the collisions would generate an enormous data stream ($\sim 30\,\text{PB/s}$), that would be unsustainable for the data acquisition, storage, and computing systems. Only $\sim 1\,\text{kHz}$ of events can be recorded by CMS and a two-level trigger system is used to select the events to be written to disk. The trigger system is composed by the Level-1 trigger (L1) and the High Level Trigger (HLT).

2.4.1 L1 trigger

The L1 trigger is a hardware trigger that performs the first event selection reducing the rate from $\sim 30\,\text{MHz}$ down to $\sim 100\,\text{kHz}$. The scheme in Fig. 2.41 summarizes the logic of the L1 trigger. The energies deposited in the ECAL, HCAL, and HF towers are read by the "regional calorimeter trigger" (RCT), that identifies electrons/photons, jets, and taus. The objects are then passed to the "global calorimeter trigger" (GCT) that sorts them in momentum and evaluates the total transverse energy and the total

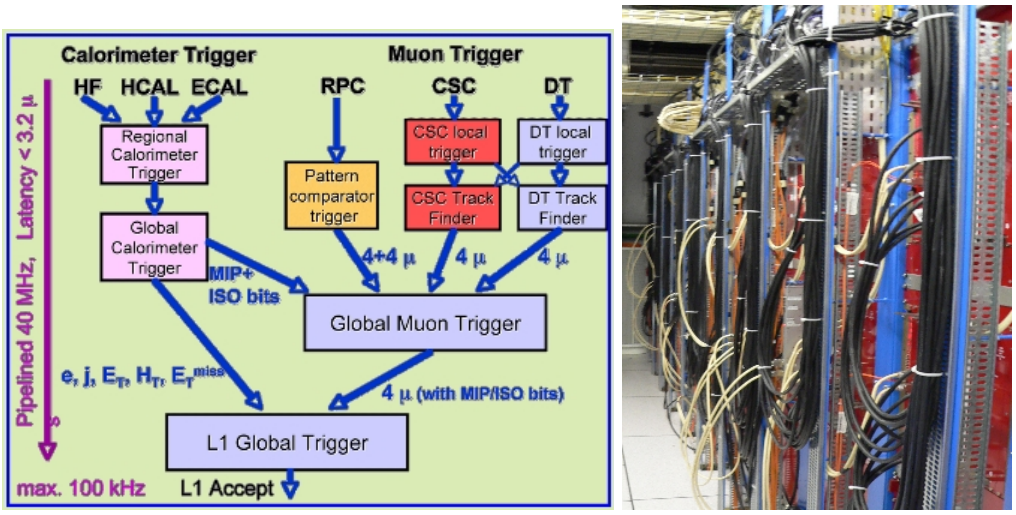


FIGURE 2.41: Scheme of the working of the L1 trigger (left) and a picture of the Regional Calorimeter Trigger (right) [66, 67].

transverse missing energy vector. The hits in the muon chambers (RPC, DT, CSC) are identified and used to build the muon track candidates. The candidates are passed to the "global muon trigger" (GMT) that defines the final muon tracks. Eventually, the GCT and the GMT are combined in the "global trigger" (GT) decision to accept or reject the event, if it matches all the requirements of at least one trigger of the L1 menu (e.g. MET $> 70\,\text{GeV}$). The final decision bit is then brought back to the sub-detectors through the Trigger and Control system (TTC) in order to start the full readout of the detectors.

Phase-I upgrade

The L1 trigger has been recently upgraded according to the CMS Phase-I upgrade plan. It includes two steps: L1 stage-1 and stage-2 upgrades [68].

- *Stage-1.* The stage-1 L1 upgrade was accomplished during the summer of 2015. It improved the calorimetric trigger by adding two new features: the identification of tau jets, and the pile-up subtraction for jets and energy sums. Minor improvements were achieved on the muon trigger: new muon chambers, higher granularity in CSC, and small changes in the track finder.

- *Stage-2.* The stage-2 L1 upgrade was completed during the technical stop before the 2016 collisions. The main changes in the calorimeter trigger were an increase of the granularity and an improvement on the pile-up subtraction. The momentum resolution of the muon trigger were significantly improved as well as the muon isolation. With the upgrade, the logic of the L1 Global Trigger is more powerful and can handle more complex topologies.

2.4.2 High Level Trigger

The HLT is a software trigger that accessing the whole information of all CMS detectors reduces the rate from $\sim$ 100 kHz to $\sim$ 1 kHz. Every event accepted by the HLT is stored on disk waiting for the offline reconstruction. The HLT is a computer farm composed of 16000 CPUs and hence the maximum average latency supported by the HLT is approximately 16000/100 kHz = 160 ms. In order to keep the HLT latency within the limit, a dedicated simplified event reconstruction is used instead of the offline reconstruction that would require tens of seconds. The maximum HLT rate is limited around $\sim$ 1 kHz. This requirement is due to the limited offline computing resources and, secondarily, to the mass storage system.

Some searches of new physics, e.g. the search for a di-jet resonance in the low mass region, need a huge amount of statistics to be competitive with the previous measurements and, therefore, they are strongly constrained by the HLT rate limit. In the attempt of going beyond it, at the CMS there are two additional special ways of data taking.

- *The data scouting.* It consists in saving only the HLT-reconstructed objects, instead of the full event. The data-scouting analyses use directly those variables and, therefore, they do not need the offline reconstruction. For these reasons, they can exceed the limit of $\sim$ 1 kHz on the HLT rate. In addition, as the size of the reduced events is only $\sim$ 10 kB instead of $\sim$ 1 MB, the data scouting does not stress the mass storage system.
- *The data parking.* As the main limit on the HLT rate comes from the offline computing system, additional events can be stored ("parked") but without reconstructing them. These events will be reconstructed opportunistically in the future during the periods when the offline computing system load is small, for instance during the next LHC Long Shutdown. In Run 2, the data parking is used by CMS together with the data scouting. In this way, in case a data-scouting analysis observes an anomalous signal it can be confirmed or excluded later using the full event information stored in the parked data.

2.4.3 An example of high level trigger path

The HLT is made up of a set of $\sim$ 500 trigger paths. Each trigger path is a sequence of software modules that reconstruct relevant physics quantities (eg. MET) with improved precision and filter on them (e.g. requiring MET $>$ 170 GeV). As soon as an event is rejected by a filter the subsequent reconstructing modules in the path are skipped. When an event passes all the filters contained in a trigger path, it is written on disk.

The HLT has two constraints: the average latency ($\lesssim$ 150 ms) and the global rate ($\lesssim$ 1 kHz). These two constraints come, respectively, from the number of CPUs available

in the HLT farm and from the computing and storage system capabilities. In order to keep low latency, the HLT reconstructs in the earliest stages the physics objects that can be quickly reconstructed, e.g. calorimetric jets, and uses them for a first loose event selection. Afterwards, a more precise and tight event selection is performed using objects reconstructed with a more performing but slower reconstruction program (e.g. particle flow, PF). In this way, the latter and slower reconstruction contributes to the average HLT latency only for the fraction of events passing the first selection.

In order to explain in some details on the functioning of an HLT trigger path, we consider as an example a trigger that requires MET $> 120\,\text{GeV}$ plus one b-tagged jet. The numbers reported in this section have been obtained using 13 TeV minimum bias simulation with pile-up 40 and 25 ns of bunch spacing.

A typical sequence of selections used in a HLT path with b tagging uses the following information: L1 seed, calorimetric objects, b tagging with regional tracking, and PF objects.

- **L1 seed** *[Input rate* $\sim 100\,\text{kHz}$*; Time* $< 1\,\text{ms}$*].* Each HLT path runs over every event accepted by the L1 trigger ($\sim 100\,\text{kHz}$). The first cut of a HLT path is the selection of a specific L1 trigger seed, i.e. a specific trigger in the L1 menu or a logical combination of them. In the example, the L1 seed used is MET $> 70\,\text{GeV}$. The cut on the L1 seed reduces the rate of a factor ~ 10.

- **L2 - Calorimetric selection** *[Input rate* $\sim 10\,\text{kHz}$*; Time* $\sim 50\,\text{ms}$*].* Since the reconstruction of the calorimetric objects (e.g. calorimetric MET) is very fast, it takes place usually just after the L1 seed selection. The resolution of these objects is worse compared to the PF objects, but they can be reconstructed in only $\sim 50\,\text{ms}$, 20 times faster than the corresponding PF objects. Calorimetric objects can be reconstructed for any event accepted by the L1 trigger. When the selection criteria are applied to calorimetric objects tracks are not available, and hence we cannot tag pile-up jets. In the example, the L2 cuts used are: calorimetric MET $> 70\,\text{GeV}$ and two calorimetric jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.6$. Usually, the L2 cuts reduce the rate of a factor 5 – 10.

- **L3 - Regional tracking** *[Input rate* $\sim 2\,\text{kHz}$*; Time* $\sim 100\,\text{ms}$*].* As tracking is a slow process, the tracker information is not used in the L1 and L2 stages. Still, tracks are essential for many objects like distinguishing between photons and electrons, identify b and τ jets, evaluate the track-based lepton isolation, and measure precisely the lepton momentum. To take advantage of the tracking information at HLT, two strategies are used: run tracking over a small fraction of events and only in the regions of interest (cones with $\Delta R = 0.4$ around the calorimetric objects). For these reasons, many triggers perform a regional tracking after the L2 cuts to evaluate the variables listed above. In the example, the regional

tracking is used to perform b tagging and it is used to reduce the rate down to $\sim 500\,\text{Hz}$.

- **PF sequence** *[Input rate* $\sim 500\,\text{Hz}$*; Time* $\sim 1000\,\text{ms}$*].* The PF sequence is the most accurate and slow reconstruction sequence available at HLT. Usually, the tightest cuts at HLT are based on the PF objects. The PF sequence used at HLT is similar to the offline version, but it uses a simplified version of tracking. The HLT PF tracking reconstructs only tracks close to leading primary vertices and with $p_T > 500\,\text{MeV}$. Then, tracks are linked to objects reconstructed with the calorimeters and the muon chambers. The PF sequence reconstructs any high level objects: jets, MET, electrons, photons, muons, taus, etc. In some HLT paths, the PF tracks are also used to perform b tagging on PF jets. In the example, the PF objects are used for the final cut MET $> 120\,\text{GeV}$ bringing the rate down to $\sim 10\,\text{Hz}$.

The input rate, time consumed, and average-per-event time consumed of each HLT step is reported in Table 2.2. Note that even if the PF sequence needs $\sim 1000\,\text{ms}$, it increases the average path time of only $\sim 5\,\text{ms}$ because it is run at $\sim 500\,\text{Hz}$.

The rate and timing plots for the MET+b-tagging trigger are shown in Fig. 2.42. In details, Fig. 2.42a shows the trigger rate as a function of the reconstructing and filtering software modules: it starts from the L1 trigger rate ($\sim 100\,\text{kHz}$) and ends with $\sim 10\,\text{Hz}$. The five steps in the plot correspond to the modules filtering on L1 seeds, calorimetric MET, calorimetric jet, b tagging, and PF MET. Fig. 2.42b reports the time per running modules and Fig. 2.42c is the respective cumulative function. Fig. 2.42d is the cumulative function of the average time spent on each module and the maximum value ($\sim 9.3\,\text{ms}$) is the average time needed to run this HLT path.

TABLE 2.2: Time and rate of the object reconstruction of a MET + b-tagging trigger running on 13 TeV minimum bias simulation with pile-up 40 and 25 ns of bunch spacing. The values change depending on simulation and trigger path used. The numbers reported are the running time, the input rate, and the average time spent per event. The last column can be calculated as the running time multiplied by the input rate and divided by the global HLT input rate ($\sim 100\,\text{kHz}$). For reference, the limit at HLT on the average global latency is $\sim 150\,\text{ms}$ and on the rate is $\sim 1\,\text{kHz}$.

Object reconstruction	$\sim$ Time (ms)	$\sim$ Rate (Hz)	$\sim$ Av. time (ms)
L1 variables	< 1	10^5	< 1
Jet/MET	50	10^4	5
Regional tracking (b tagging)	100	2000	2
Particle flow	1000	500	5
Data storing	–	10	–

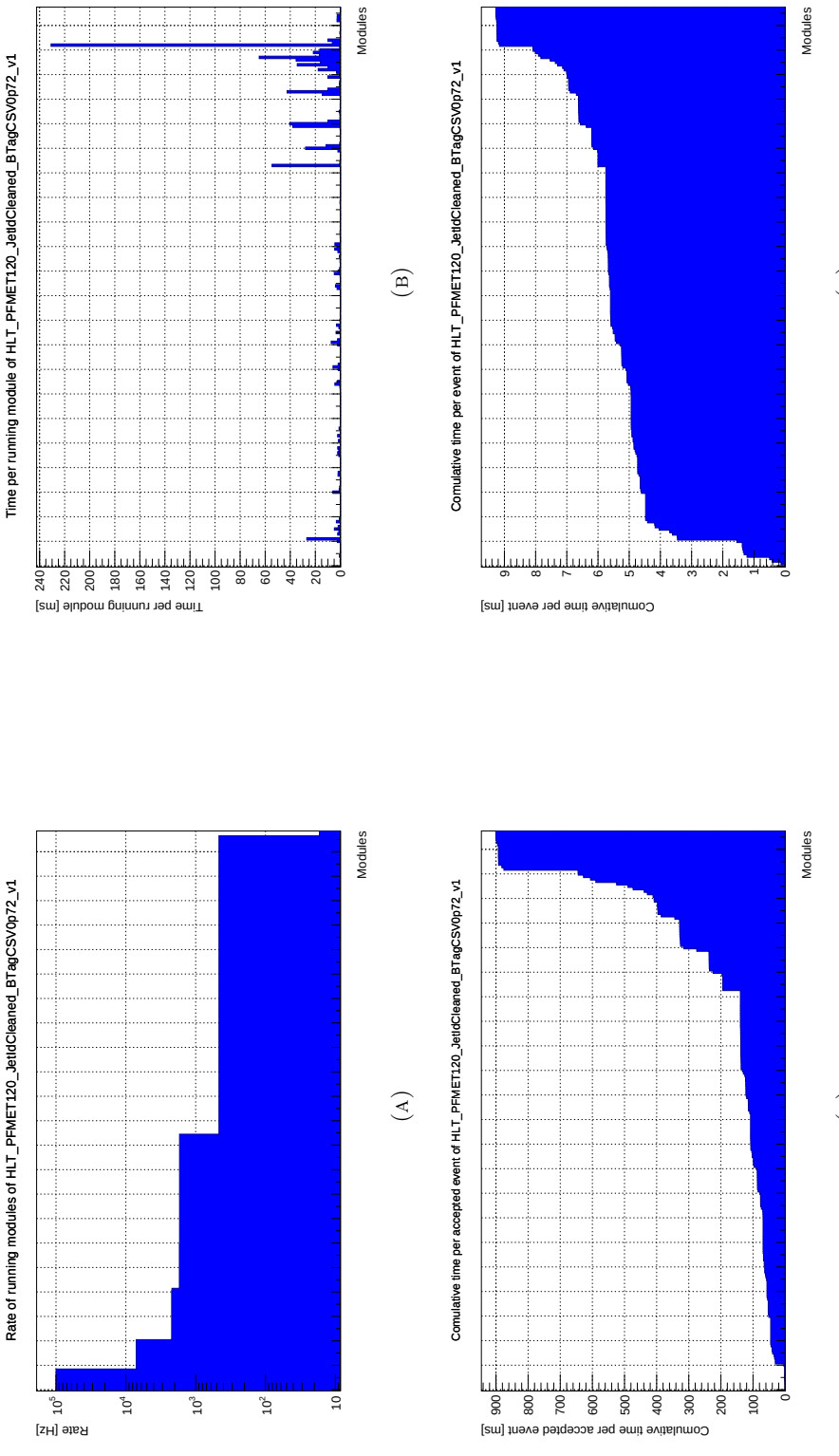


FIGURE 2.42: The timing of a MET + b-tagging trigger running on 13 TeV minimum bias simulation with pile-up 40 and 25 ns of bunch spacing. Fig. (A) shows the rate of the running modules: the five steps correspond to the cuts on L1 seeds, calorimetric MET, calorimetric jet, b tagging, and PF MET. The time spent by each running module is reported in Fig. (B) and the corresponding cumulative function is shown in Fig. (C). Fig. (D) is the cumulative function of the per-event-average time spent by the modules.

Chapter 3

Trigger improvements for hadronic $H(bb)$ channels

This chapter describes several improvements to the CMS HLT that I developed for the Run-2 data taking during the preparation of this thesis. These developments are motivated by the study of the Higgs decays to a pair to b jets, especially $Z(\nu\nu)H(bb)$, and include the improvement of tracking, b-tagging and MET definitions at HLT. However, the results of this work are much more general. In particular, exploiting these new algorithms it is possible to reject efficiently pile-up jets at an early stage of the HLT with large benefit for many CMS analyses.

In this chapter, after a discussion of the challenges of HLT at Run 2, I will describe the fast pixel tracking that can be run on every event accepted by the L1 trigger ($\sim 100\,\text{kHz}$) allowing identification of pile-up jets. This fast tracking is also used to improve the timing of the online b-tagging algorithm developed firstly for the search for $H(bb)$. The last part of the chapter describes the improvements developed on the reconstruction of the MET variable at HLT and the preparation of triggers dedicated to the search for the $H(bb)$ in events without charged leptons in the final states.

3.1 The challenges of Run 2

In 2015 LHC started Run 2 with the first proton-proton collisions at the center-of-mass energy of 13 TeV and bunch spacing of $\sim 7.5\,\text{m}$ (25 ns). The maximum instantaneous luminosity achieved in 2015 was about $5 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$. In the next few years, LHC is expected to increase further the instantaneous luminosity up to $1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, reaching about 40 pile-up collisions.

This high luminosity scenario, compared to Run 1, has strong consequences for the trigger system.

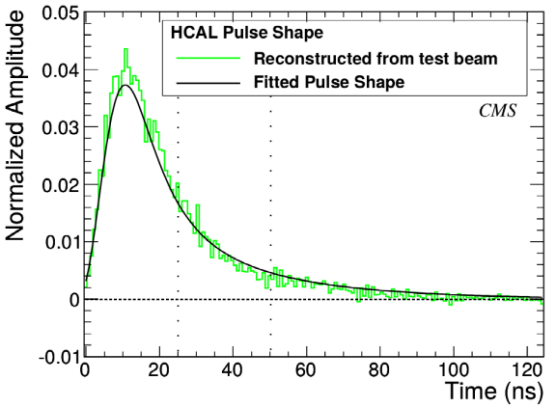


FIGURE 3.1: Average pulse shape of HCAL from test-beam data using 300 GeV pions. The vertical lines correspond to 25 ns (Run 2) and 50 ns (Run 1). In Run 2, large out-of-time pile-up effects are expected [69].

- *Higher pile-up.* The mean number of proton-proton collisions per bunch crossing is expected to be about 40 in the high luminosity Run 2, whereas it was about 20 during Run 1. Pile-up collisions deposit additional energy in calorimeters, degrading the resolution of all calorimeter-related variables like jet energy, MET, and lepton isolation. Pile-up can also produce additional fake jets, commonly called pile-up jets. These effects are more detrimental in the L1 trigger and in the first stages of HLT where tracks are not yet available and pile-up objects cannot be distinguished from signal objects. In addition, the tracking time increases because of extra tracks from the pile-up collisions.

- *Higher energy.* The center-of-mass energy of 13 TeV, compared to the 8-TeV scenario, shifts the spectrum of particles energy to higher values and increases the total number of particles produced in the collision, raising the trigger rates for fixed thresholds. Moreover, the mean energy released by each pile-up collision is larger, enhancing the pile-up effects described above.

- *Smaller bunch spacing.* The reduction of the bunch spacing reduces the pile-up at fixed luminosity. However, using 25 ns of bunch spacing a new effect occurs: the HCAL energy measurements are contaminated by the previous bunch crossing. Figure 3.1 shows a typical pulse of HCAL readout. Only 68% of the pulse is contained within 25 ns, and hence the residual part contaminates the measurements of the following bunch-crossing. A similar issue, called out-of-time pile-up, occurs in the ECAL as well as in the tracker, however the effect is much smaller since these two detectors have a faster readout. These effects were much reduced during Run 1 because LHC used a bunch spacing of 50 ns.

- *Higher luminosity.* In Run 2, the instantaneous luminosity will be about doubled compared to Run 1. Therefore, on top of the other effects, the $\mathcal{L} \cdot \sigma$ of all process, including the background, are doubled and this will increase the rate by factor two.

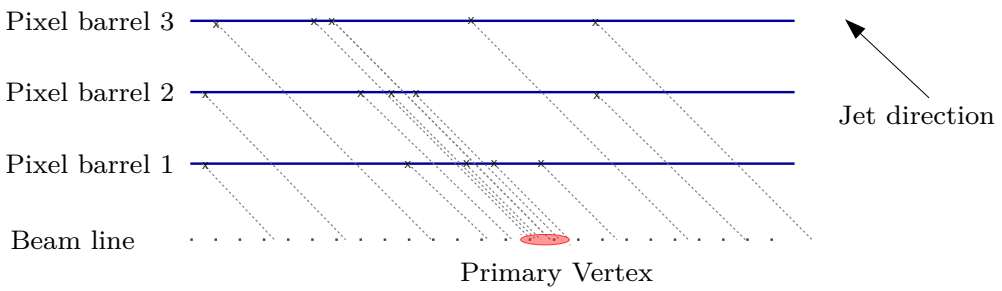


FIGURE 3.2: Sketch of the algorithm used for the fast localization of the primary vertex. The blue lines represent the pixel barrels in the $r - z$ plane, the "X" are the pixel hits and the dotted line is the beam line. The algorithm projects back, along the jet direction, the pixel hits to the beam line. The region with high concentration of projected pixel hits defines the primary vertex.

These effects are more evident in the first stages of the trigger when tracks are not available (L1, first stages of HLT). To cope with these effects, I have developed new tools that are described in the next sections. The idea is to introduce a fast pixel tracking that can be run on every event accepted by the L1 trigger ($\sim 100\,\text{kHz}$). This novel tool is very fast and is used both to reject pile-up jets in the early stage of HLT and to make the online b tagging faster.

These new tools, along with an improvement of the online b tagging and MET reconstruction, has been used to prepare a set of triggers to collect data for all hadronic $H(bb)$ analyses in Run 2.

3.2 Fast pixel tracking

3.2.1 Fast primary vertex

The first step of the fast pixel tracking is a rough localization of the primary vertex. The primary vertex position in the transverse plane is well defined by the beam spot, with a precision of about $20\,\mu\text{m}$. The localization along the beam direction (z) is performed by an algorithm that, without using tracks, identifies the primary vertex within an uncertainty of about 3 mm. The idea is that tracks contained in a jet have roughly the same direction of the jet, that is pretty well measured with the calorimeters. Therefore, the primary vertex can be found projecting back, along the jet direction, the pixel hits associated to the jet. Figure 3.2 illustrates this idea.

The presence of additional pixel hits, produced by tracks emerging from pile-up vertices or from other jets, pollutes the fast vertexing algorithm. To mitigate this effect, pixel hits are selected according to the length and width of their pixel cluster, and their φ coordinate. The incident angle of a high p_T track with the pixel barrel detector is about a right angle on the transverse plane and is related to the jet azimuthal angle θ on the

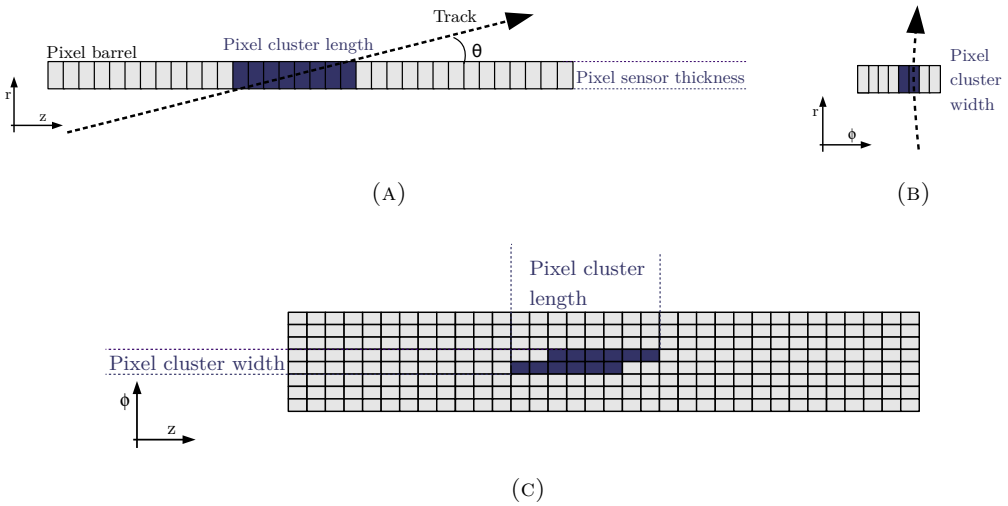


FIGURE 3.3: Figure shows the shape of a pixel cluster along three projections: $r-z$ (A), $r-\varphi$ (B), and $z-\varphi$ (C). As shown in Fig. (A), the pixel cluster length, in the longitudinal direction, depends only on the azimuthal track angle θ. The cluster length can be calculated as (pixel thickness)$/\tan\theta$. Therefore the mean number of pixels in the cluster is given by $1+$ (cluster length)/(pixel thickness). Fig. (B) shows that high p_T tracks are roughly orthogonal to the pixel module. The cluster width of high p_T tracks is of two pixels because of the Lorentz angle effect. Fig. (C) is the 2D shape of a pixel cluster.

longitudinal plane. As shown in Fig.3.3, the shape of the pixel clusters of high p_T tracks is expected to have a width of two pixels on the traverse plane, due to the Lorentz drift (see Section 2.2.2), and a length that depends on the jet azimuthal angle on the longitudinal plane:

$$\text{cluster length} = 1 + \frac{\text{pixel thickness}/\tan\theta}{\text{pixel length}} = 1 + \frac{1.9}{\tan\theta} \tag{3.1}$$

where the cluster length is the mean number of pixels along η direction, the pixel sensor thickness is $285\,\mu$m, and the pixel length is $150\,\mu$m. Figure 3.4 is a 2D histogram that shows the linear dependence between the cluster length and $1/\tan\theta$ of the jet, as predicted by (3.1). The overall effects of the hits selection, based on the φ coordinate and cluster shape, are shown in Figure 3.5.

Briefly, the fast primary vertex can be localized using the following steps:

1. jet calorimetric reconstruction;
2. pixel hits reconstruction;
3. the hits with φ coordinate and cluster shape compatible with the four leading jets having $p_T > 30\,\text{GeV}$ are selected;
4. the selected hits are projected back to the beam line using the jet η coordinate;

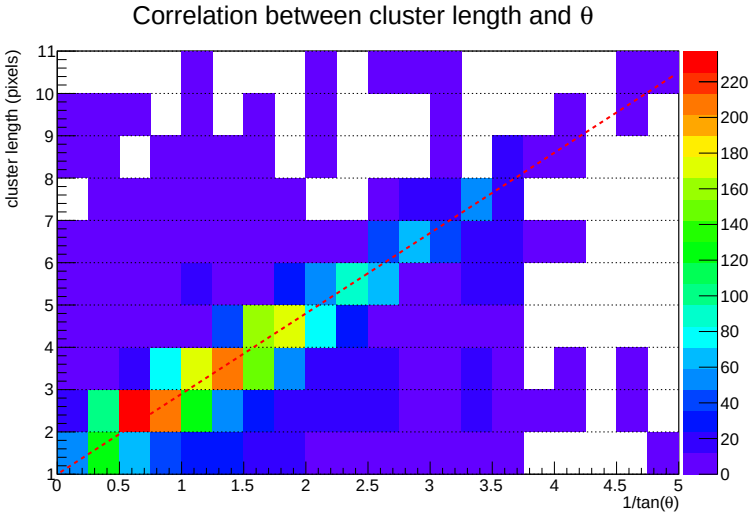


FIGURE 3.4: Pixel cluster length in the longitudinal direction as a function of $1/\tan\theta$ of the jet. The 2D histogram has been realized using simulated events and selecting, for each jet, the pixel clusters with $\Delta\varphi(\text{jet}, \text{cluster}) < 0.2$, and projected-$z$ within 0.5 cm from the primary vertex. The dashed line shows the expected dependency, reported in (3.1). The histogram shows that the clusters lengths are compatible with the expectation. The tails of the distribution are due to low momentum tracks originating in pile-up vertices. The plot has been produced with $t\bar{t}$ simulation with pile-up 60 and $\sqrt{s} = 8$ TeV.

5. the peak in z coordinate binned distribution of the projected hits is identified as the fast primary vertex.

We developed this algorithm in 2012 and it was successfully used during the data-taking. The algorithm has been improved for Run 2 with the following new features.

1. The acceptance of jets used in the fast vertexing has been extended from $|\eta| < 1.6$ (barrel only) to $|\eta| < 2.4$ (barrel + end-cap). The algorithm is slightly different for the end-cap, as they have a different orientation of the pixel sensors. In particular, here the cluster length is proportional to $\tan(\theta)$, instead of $1/\tan(\theta)$.

2. Each hit is weighted with a rough estimation of the likelihood of the hit of belonging to the jet tracks. This likelihood has been estimated using the following variables: $\Delta\varphi(\text{jet}, \text{hit})$, cluster width and length, and cluster charge. Figure 3.6 shows the distributions of these variables in $t\bar{t}$ simulations, both for the hits related to the signal jet tracks and for the pile-up jet tracks. Hits are also weighted depending on their pixel layer, as the layer closest to the interaction point gives the highest resolution. Each hit is weighted also with the jet p_T since high momentum jets contain more signal hits and hence a lower relative pile-up hit contamination.

3. The method used to find the peak in the projected z hits distribution has been improved. In our first implementation, it was a search of the 2-cm range containing the maximum number of projected clusters. The new method, instead, iterates

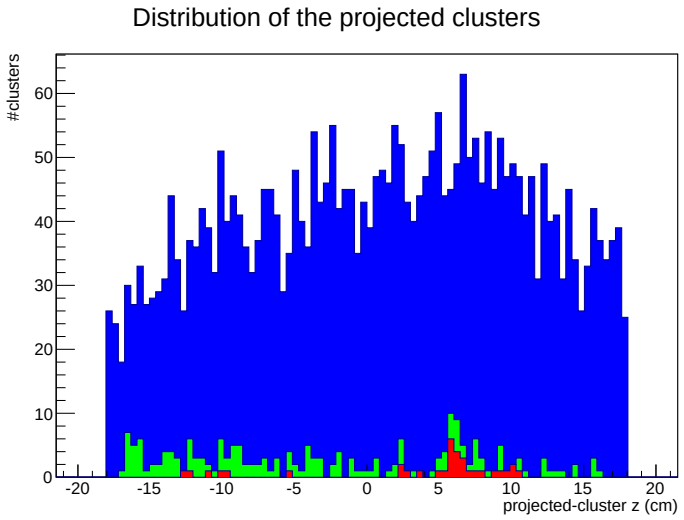


FIGURE 3.5: z distribution of the projected hits in one simulated $t\bar{t}$ event with $\sqrt{s} = 8\,\text{TeV}$ and pile-up 60. The variable is the z coordinate of the pixel hits projected back to the beam line, along the jet direction. The distributions has been obtained for: inclusive hits (blue), hits with $\Delta\varphi(\text{jet}, \text{hits}) < 0.2$ (green), and hits with cluster shape compatible with Eq. (3.1) and $\Delta\varphi(\text{jet}, \text{hit}) < 0.2$ (red). The peak around $z = 6\,\text{cm}$ is identified as the fast primary vertex. The histogram shows that after each selection the primary-vertex peak appears more clearly. In this event, the generated primary vertex was at $z_{\text{true}} = 6.51\,\text{cm}$.

the search more times, using smaller ranges. In addition, in order to improve the resolution, only high-weighted hits are considered in the last step, because they have the best resolution.

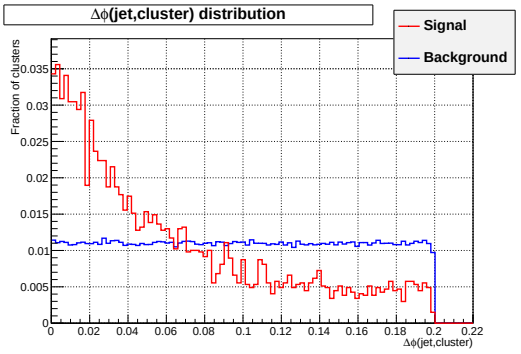
The algorithm performance has improved: the efficiency of finding the correct primary vertex has been increased, as shown in Figure 3.7, and also the resolution has improved, as shown in Figure 3.8. Table 3.1 shows the efficiency of finding the primary vertex for the new algorithm as a function of the number of jets with $p_T > 30\,\text{GeV}$ used in the algorithm.

The efficiency of reconstructing the primary vertex within 1.5 cm is about 90%. In the next section will be shown that the knowledge of the position of the primary vertex permits a reduction of the time needed for tracking with pixel of about 80%.

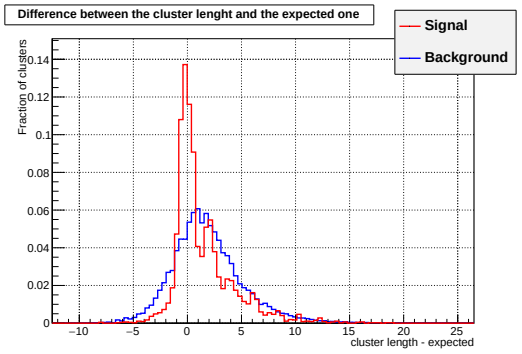
3.2.2 Pixel tracking

The first step of tracking is the reconstruction of the tracks in the pixel detector (pixel tracking). The pixel tracking requires about 100 ms for reconstructing all tracks with $p_T > 0.9\,\text{GeV}$ in 13 TeV minimum bias simulation with pile-up 40. This high latency makes impossible to run the pixel tracking for all events accepted by the L1 trigger.

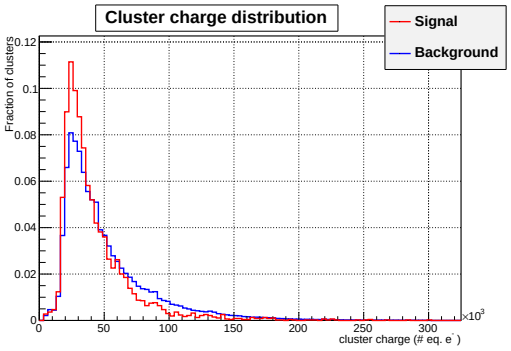
We reduced the latency of the pixel tracking in two ways. The first one is reconstructing only tracks originating within 1.5 cm from the fast primary vertex. This strategy was



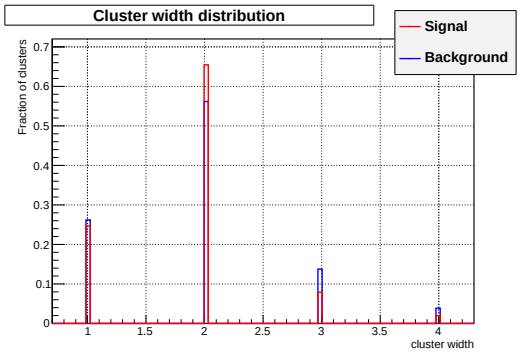
(A) $\Delta\varphi$(jet, hit).



(B) Difference between the cluster length and the one expected (Eq. 3.1).

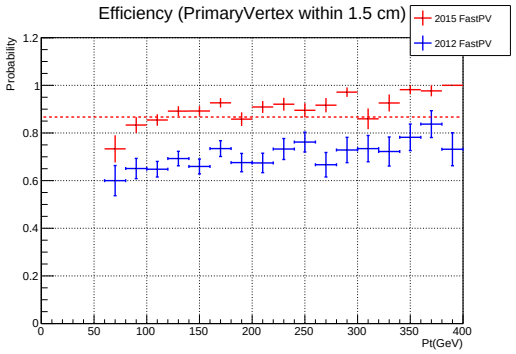


(C) Cluster charge.

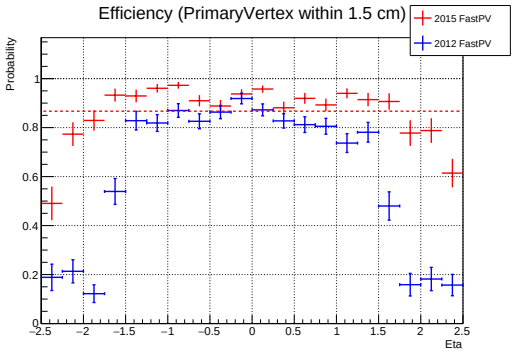


(D) Cluster width.

FIGURE 3.6: Variables used to evaluate the likelihood of the hits of belonging to the jet tracks. Each variable is plotted for the clusters produced by the pile-up (blue) and signal jet (red) tracks. The plot has been produced using a $t\bar{t}$ simulation with pile-up 60 and $\sqrt{s} = 8\,\text{TeV}$.



(A) Efficiency as a function of p_T.



(B) Efficiency as a function of η.

FIGURE 3.7: Efficiency of finding the fast primary vertex within 1.5 cm with the improved (red) and default (blue) algorithm. In general, the algorithm uses the four leading jets with $p_T > 30\,\text{GeV}$ to find the primary vertex. Here, the primary vertex has been reconstructed using only the leading jet in order to monitor the efficiency as a function of the jet p_T (A) and η (B). The plot has been produced using a $t\bar{t}$ simulation with pile-up 40 and $\sqrt{s} = 13\,\text{TeV}$, considering only events with a leading jet with $|\eta| < 2.4$. The dashed line corresponds to the mean efficiency after this preselection.

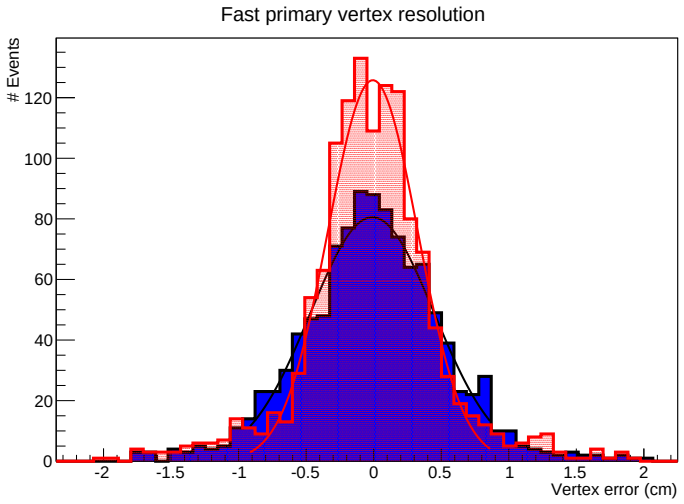


FIGURE 3.8: The difference between the fast primary vertex and the true vertex using the 2012 (blue) and 2015 (red) algorithm. The distributions are fitted with Gaussian having $\sigma_{\text{baseline}} = 4.6\,\text{mm}$ and $\sigma_{\text{improved}} = 3.3\,\text{mm}$. The plot has been produced using $Z(\nu\nu)H(bb)$ simulation with pile-up 40 and $\sqrt{s} = 13\,\text{TeV}$.

already used in 2012, and actually it reduced the time of pixel tracking down to $\sim 20\,\text{ms}$. The second one, introduced in Run 2, is the regional tracking: only tracks compatible with both the fast primary vertex and one of the four leading jets with $p_T > 30\,\text{GeV}$ are reconstructed. This new method reduces further the timing of about a factor five, making affordable running the pixel tracking for any event accepted by the L1 trigger ($\sim 100\,\text{kHz}$).

Table 3.2 reports the average latency of the different tracking scenarios. The regional tracking with the primary vertex constraint is the faster option and, as it will be explained in the following section, it improves the efficiency of finding the correct primary vertex using the reconstructed tracks for the $Z(\nu\nu)H(bb)$ signal.

TABLE 3.1: The efficiency of the fast vertexing to find the primary vertex within 1.5 cm from the true position for $t\bar{t}$ simulated events with pile-up 40, as a function of the number of jets with $p_T > 30\,\text{GeV}$ used in the algorithm. Currently, four jets are used in the fast primary vertex reconstruction.

Number of jets	Efficiency (%)
1	86.7%
2	93.2%
3	94.7%
4	95.4%

3.2.3 Recovery of fast vertexing failure

The fast vertexing is useful to reduce the trigger timing, but its efficiency is only about 90%. In 10% of the cases, the algorithm delivers a wrong primary vertex causing a loss of tracking efficiency. This issue can be recovered after the pixel tracking. In case the fast primary vertex reconstructs the wrong primary vertex, just few or no tracks are reconstructed in the jets because of the primary vertex constraint. In Run 1, the fast primary vertex failure was tagged using the ratio R_j, evaluated on the two leading jet, defined as:

$$R_j = \frac{\sum_{\mathrm{i}} p^T_{i,j}}{E^T_j}, \tag{3.2}$$

where $p^T_{i,j}$ is the momentum of the i track associated with the jet j, and E^T_j is the calorimetric transverse energy of the jet j. In case a jet had $R_j < 10\%$, the pixel tracking without vertex constraint was run around that jet, in order to find the correct position of the primary vertex. Only two jets were used in this step, as tracking without vertex constraint is slow and two jets are sufficient to find the primary vertex with more than 95% of efficiency.

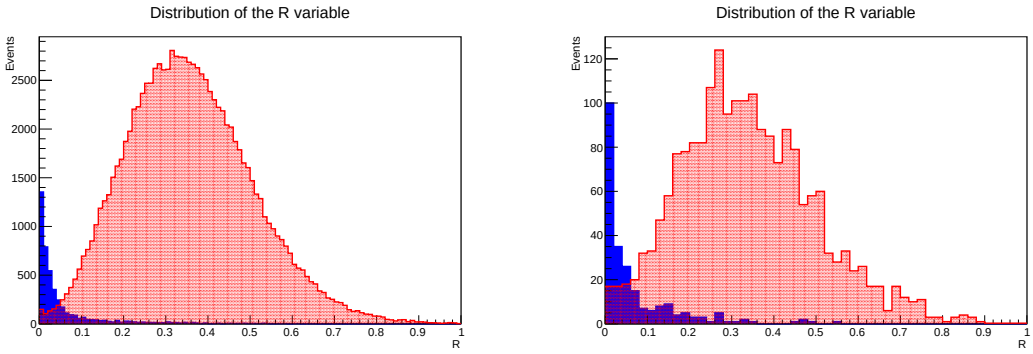
In Run 2 a new variable is used. When the fast vertexing fails, the two leading jets are expected to contain no tracks and hence it is more performing to use only one discriminator, defined as:

$$R = \frac{\sum_{\mathrm{j}} \sum_{\mathrm{i}} p^T_{i,j}}{\sum_{\mathrm{j}} E^T_j} \qquad \text{with } j \in [1,2]. \tag{3.3}$$

In case $R < 10\%$, the pixel tracking without vertex constraint is run around both jets. The discriminant R is evaluated considering only tracks fitted with $\chi^2 < 20$, in order to remove the fake tracks. As the pixel tracker is not able to measure precisely the

TABLE 3.2: Comparison of timing for different pixel tracking scenario. "Vertex constraint" means that only tracks within 1.5 cm from the fast primary vertex are reconstructed. "Regional tracking" means that only tracks with $\Delta R < 0.4$ from the four leading jets are reconstructed. In order to compare the performances, the efficiencies of finding the primary vertex within 1 mm using tracks is also reported. The efficiency has been evaluated with a Z($\nu\nu$)H(bb) simulation having MET > 130 GeV and the timing with a minimum bias simulation. Both samples have been generated with $\sqrt{s} = 13$ TeV and pile-up 40.

Year	Pixel tracking	Time	Primary vertex < 1 mm
(2011)	Full tracking	119.4 ms	84.9%
(2012)	Vertex constraint	21.4 ms	90.5%
–	Regional tracking	8.0 ms	91.7%
(2015)	V. constraint & reg. tracking	3.3 ms	90.7%



(A) $t\bar{t}$ simulation. (B) Z($\nu\nu$)H(bb) simulation.

FIGURE 3.9: Distributions of the ratio R, as defined in (3.3), for inclusive $t\bar{t}$ (A) and Z($\nu\nu$)H(bb) (B) simulations with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. This variable is used to identify the events where the fast vertexing fails. The histograms have been produced for events where the fast vertex was within (red) or outside (blue) 1.5 cm from the generated primary vertex.

momentum of high p_T particles, the maximum value of pixel track p^T considered is 20 GeV. Figure 3.9 shows the distribution of R for $t\bar{t}$ and Z($\nu\nu$)H(bb) simulations with pile-up 40. As expected, the variable R is small when the fast vertexing fails. Table 3.3 reports a comparison between the two fast vertexing recovery algorithms. The new method is better as the efficiency of finding the correct primary vertex has improved, running the recovery less often. The average running time of the recovery is very small ($\sim 0.3\,\text{ms}$), as it is run only in a small fraction of events, and therefore it can be run - when needed - for each event accepted by the L1 trigger.

TABLE 3.3: Comparison between the improved and default fast vertexing recovery algorithm. The third column (Recovery run) reports the fraction of events on which the recovery is run (low is better). "No recovery" is the efficiency of finding the primary vertex without the recovery step. The fourth column shows the percentage of events that have the final primary vertex, reconstructed with pixel tracks, within 1 mm (high is better). In Z($\nu\nu$)H(bb) the fraction is lower than in $t\bar{t}$ events because of the smaller jet multiplicity and of the softer p_T spectrum of jets. The simulations have been generated with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40.

Sample	Algorithm	Recovery run	Final primary vertex $< 1\,\text{mm}$
Z($\nu\nu$)H(bb)	Improved	12.9%	89.1%
...	Baseline	22.4%	89.0%
...	No recovery	–	85.5%
$t\bar{t}$	Improved	6.6%	98.6%
...	Baseline	19.9%	97.9%
...	No recovery	–	95.2%

3.2.4 Pixel vertexing

In the large majority of the cases, when the fast vertexing plus recovery does not fail, primary vertices are fitted again with much better resolution ($\sim$ 40 μm) using the pixel tracks reconstructed within 1.5 cm from the fast primary vertex.

Due to the large pile-up, it may happen that not all pixel tracks converge in one vertex and more than one reconstructed vertex are present in the event. In Run 1, the primary vertex choice was based on the following variable:

$$w = \sum_i \begin{cases} \left(p_i^T\right)^2 & \text{if } p_i^T < 20\,\text{GeV}, \\ (20\,\text{GeV})^2 & \text{if } p_i^T \geq 20\,\text{GeV}. \end{cases}$$

where the i-index runs over the tracks associated to the vertex. The saturation point at 20 GeV was used to take into account the momentum resolution of high p_T pixel tracks.

The wrong primary vertex was selected in about 10% of Z($\nu\nu$)H(bb) simulated events. The main causes were:

1. pile-up tracks, that increase the weight of the pile-up vertices;
2. fake tracks, that have random directions;
3. split of the potentially high-weighted primary vertex in two low-weighted vertices that are close.

These causes of inefficiency have been addressed for Run 2:

1. the amount of pile-up tracks has been reduced, thanks to the regional pixel tracking (Section 3.2.2);
2. the fake tracks have been removed, selecting only tracks fitted with a low χ^2;
3. the vertex split has been solved by merging in one single vertex those vertices that are closer than 70 μm.

The improvements obtained with these changes are reported in Tab. 3.4.

3.2.5 Overall results

The overall result achieved by the fast pixel tracking sequence (i.e. fast primary vertex, pixel regional tracking, and pixel vertexing) compared to the standard full pixel tracking is the following. For Z($\nu\nu$)H(bb), the efficiency to find the primary vertex within 1 mm has increased from 84.9% to 91.8% and the time needed for the tracking and vertexing

has drastically decreased from 120.4 ms to 6.2 ms (from 164 ms to 23 ms, including the pixel cluster reconstruction).

As the average latency limit of the HLT is about 160 ms, the fast tracking can be potentially run for all events accepted by the L1 trigger ($\sim$ 100 kHz) and used to tag pile-up jets (Sec. 3.3).

The fast tracking is also used to speed up the online b tagging (Sec. 3.4). However, the accurate b tagging cannot be run at $\sim$ 100 kHz because it needs more precise and slow tracking exploiting the outer tracker.

3.3 Fast pile-up identification with pixel tracks

Pile-up is one of the most challenging problems of Run 2. Pile-up degrades the resolution of all hadronic variables, like jets momentum and MET. In the offline reconstruction, pile-up effects are mitigated with the Particle Flow (PF) reconstruction removing all charged objects associated to the pile-up vertices. At HLT, PF is too slow to be run for all events, as it takes about one second per event. This section describes a novel fast pile-up identification, that can be run for every event accepted by the L1 trigger ($\sim$ 100 kHz).

The fast pile-up identification is a new tool that exploits the fast pixel tracking to reject pile-up jet. The fast tracking, described in Section 3.2, allows to find the primary vertex with high resolution ($<$ 1 mm) in only $\sim$ 20 ms, reconstructing only tracks close to the four leading jet directions and compatible with the fast primary vertex. Once the primary vertex is localized with high resolution, the regional pixel tracking can be extended to all jets: in Z($\nu\nu$)H(bb) sample, the regional pixel tracking performed around all jets with $p_T > 20$ GeV with a tight vertex constraint takes only $\sim$ 5 ms.

Pile-up jets have small correlation with pixel tracks originating from the reconstructed primary vertex. A 3D distance track-jet is computed for each track and the jet axis,

TABLE 3.4: Fraction of events that have the primary vertex, reconstructed with pixel tracks, within 1 mm of error, for Z($\nu\nu$)H(bb) and $t\bar{t}$ simulation with $\sqrt{s} = 13$ TeV and pile-up 40. The improvements have been achieved with the changes described in Sect. 3.2.4.

Sample	Algorithm	P. vertex $<$ 1 mm (%)
Z($\nu\nu$)H(bb)	Improved	91.8%
...	Baseline	89.1%
$t\bar{t}$	Improved	99.1%
...	Baseline	98.6%

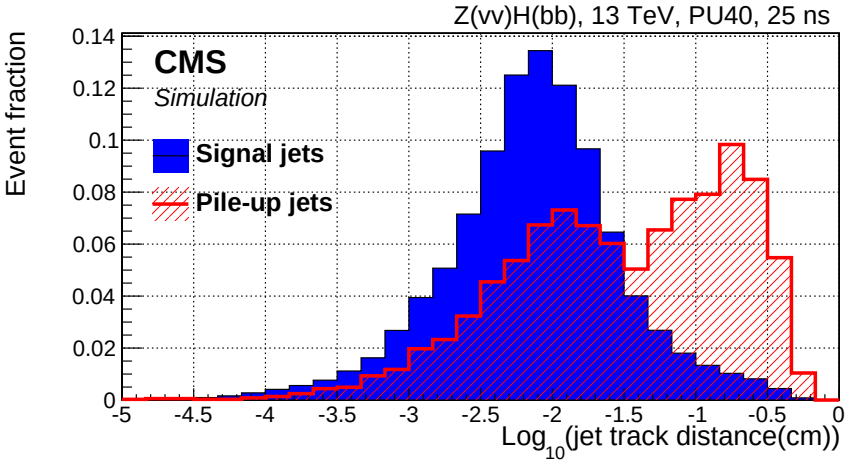


FIGURE 3.10: Minimum jet-track distance distribution for signal jet (blue) and pile-up jet (red) in Z($\nu\nu$)H(bb) sample with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. Figure shows that about 92% of tracks in signal jets are contained in 400 μm, whereas tracks in pile-up jet have a larger distance. Pile-up jets may also contain tracks with a small distance: these tracks are produced by the underlying event of the high p_T collision and have a direction close to the pile-up jet. The plots have been produced considering only tracks originating with $\Delta z < 0.5\,\text{cm}$ from the primary vertex.

defined as a straight line originating from the primary vertex and pointing towards the calorimetric jet. Figure 3.10 shows the distribution of this variable, for signal and pile-up jets. As expected, the jet-track distances in signal jets are typically around 100 μm and in pile-up jets $\gtrsim 1\,\text{mm}$.

The fraction of jet momentum carried by the pixel tracks originating from the primary vertex is expected to be small for pile-up jets and large for signal jets. Considering only tracks whose 3D distance with respect to the jet axis is smaller than 400 μm we build a pile-up discriminant with the ratio $\frac{\sum_{\text{tracks}} p_T}{p_T(\text{jet})}$, similar to the R_j variable defined in Eq. (3.2). The distribution of the pile-up discriminant is shown in Figure 3.11, demonstrating that it is actually able to separate pile-up from signal jets.

Jets are tagged as a signal if they have at least the 5% of jet momentum carried by pixel tracks. In order to assure high efficiency for the high energetic jets, the jets containing at least 2 GeV in pixel tracks momentum are in any case tagged as signal jets. The performance of the fast pile-up identification is shown in Figure 3.12 where the probabilities to tag as "no pile-up" both signal and pile-up jets are reported as a function of p_T and η. These plots show that in the central region the efficiency of tagging signal jets correctly is above 95% and the fake rate due to wrongly tag pile-up jet is below 30%.

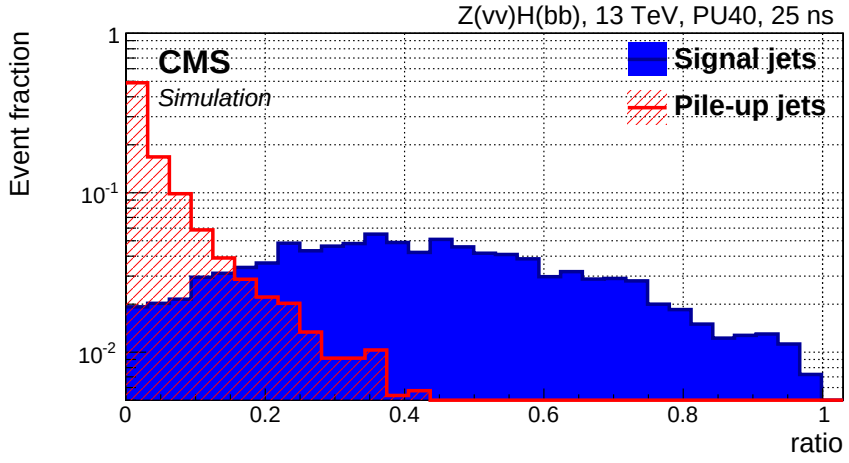
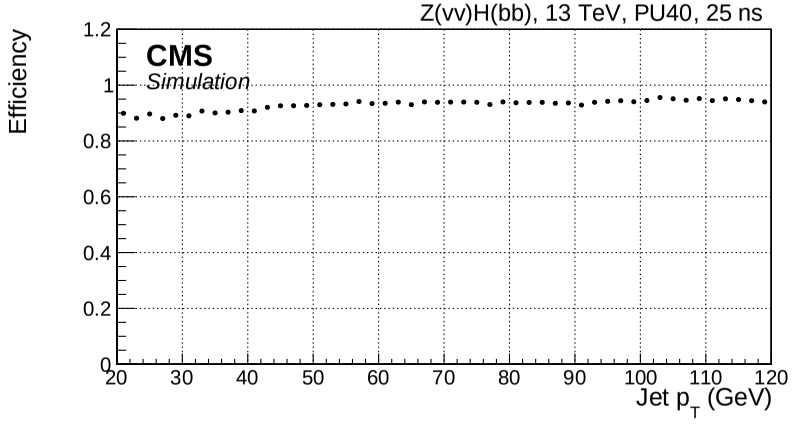
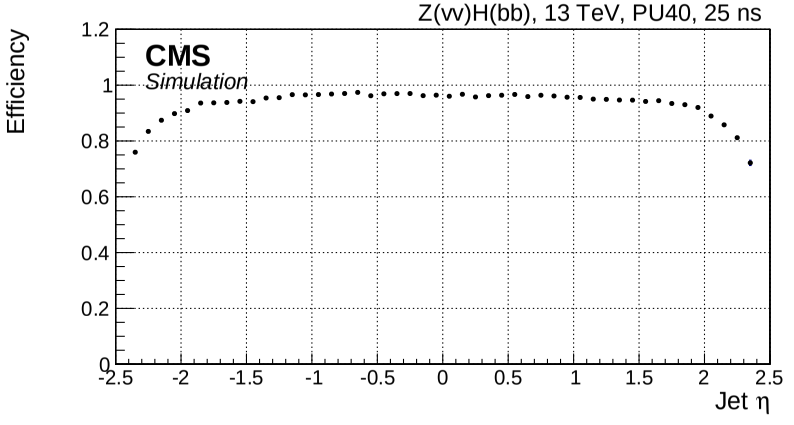


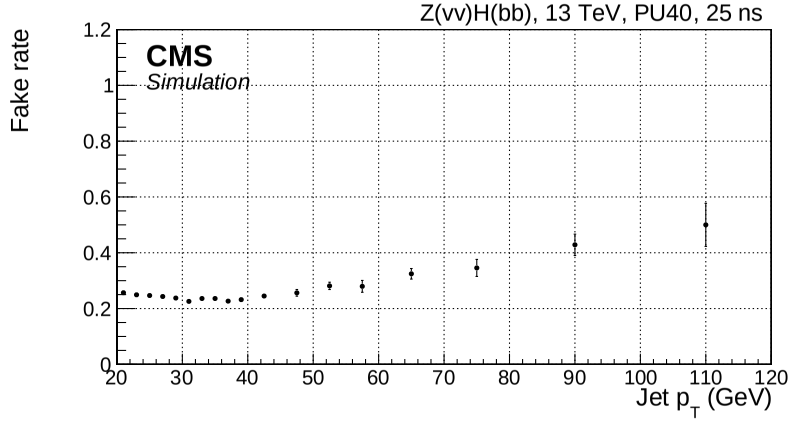
FIGURE 3.11: The ratio between jet-tracks and jet momenta, for pile-up (red) and signal jets (blue). The variable has been evaluated for each calorimetric jet with $p_T > 20\,\text{GeV}$ in Z($\nu\nu$)H(bb) simulation ($\sqrt{s} = 13\,\text{TeV}$, pile-up 40). Jets matching with $\Delta R < 0.2$ the simulated jet are considered as signal jets, whereas the jets with $\Delta R > 0.4$ as pile-up jets. As expected, pile-up jets have the ratio close to zero, while signal jets have higher values.



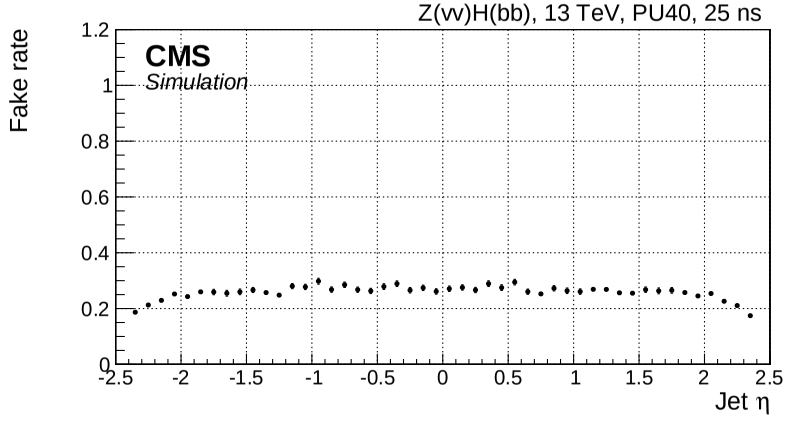
(A) Efficiency as a function of p_T.



(B) Efficiency as a function of η.



(C) Fake rate as a function of p_T.



(D) Fake rate as a function of η.

FIGURE 3.12: Probability to tag as "no pile-up" signals jets (A,B) and pile-up jets (C,D), as function of p_T (A,C) and η (B,D). The plots have been realized for jets with $|\eta| < 2.4$ and $p_T > 20\,\text{GeV}$ using VBF $H \to bb$ simulation with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40.

3.3.1 Applications of the fast pile-up identification

The fast pile-up jet identification allows to reject pile-up jet in the early stages of HLT. A direct application of this tool is to cut on the calorimetric jets after the pile-up cleaning: instead of requiring four jets with $p_T > 45\,\text{GeV}$ we can require four no-pile-up jets with $p_T > 45\,\text{GeV}$, reducing the trigger rate. The fast pile-up jet identification can be applied as well to any jet-based variable, like the scalar sum of jet p_T (HT) and the missing transverse energy evaluated using jets (MHT). Figure 3.13 compares the resolution of the calorimetric MHT and HT in a simulation with pile-up 40, before and after the fast pile-up jet cleaning. In both cases, the effects of the pile-up cleaning are significant: it improves the resolution of 25 – 30 % and largely reduces the bias in both variables.

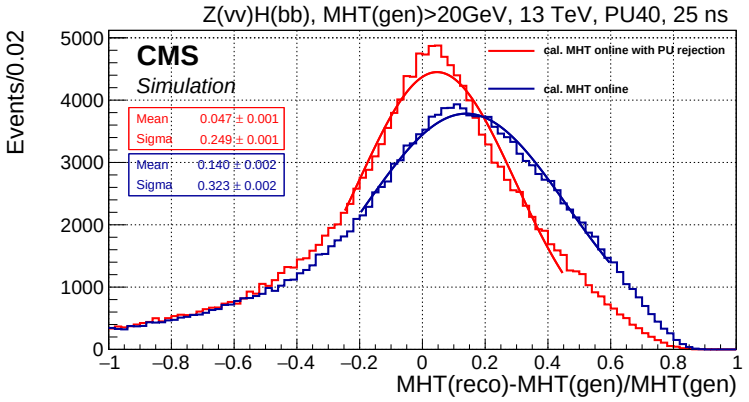
3.4 The b tagging

HLT b tagging was successfully used in Run 1. During the 7 TeV data-taking, it was performed measuring the largest impact parameter among the tracks associated with the jet. In 2012, I improved the online b tagging porting the CSV algorithm (Sect. 2.3.8) at HLT. The triggers with online b tagging allowed to improve the sensitivity of searches like Z($\nu\nu$)H(bb) [27], VBF H(bb) [29], Z(bb)H(inv.) [70], and search for resonance $X \rightarrow$ H(bb)H(bb) [71, 72].

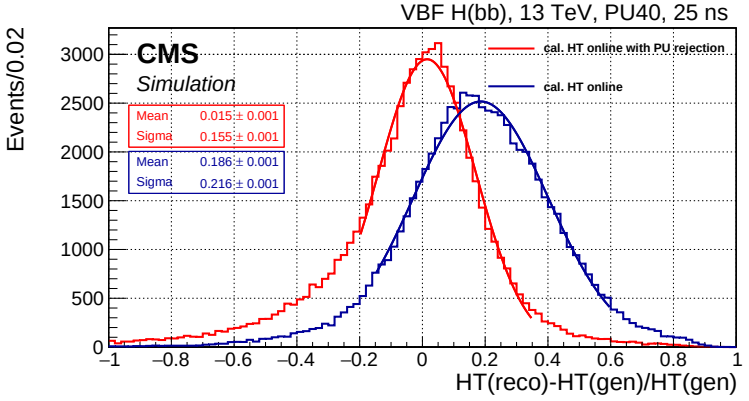
The first step of the online b tagging is the reconstruction of the pixel tracks and of the primary vertex, and this has been largely speeded up with the fast tracking described in Sect. 3.2.2. In order to have a performing b tagging, the accuracy of the impact parameter of the tracks with respect to the primary vertex must be improved using the information of the outer tracker. This improves the spatial resolution of the reconstructed tracks exploiting the full tracker lever arm in the longitudinal plane and the measurement of the momentum in the transverse plane. This step is very time-consuming, hence only tracks having a direction close to the eight leading jets with $p_T > 30\,\text{GeV}$ and originating near the primary vertex are reconstructed. The constraint on the primary vertex is quite loose (5 mm) in order to reconstruct also the tracks originating from displaced secondary vertex.

In Run 1, the seeds used in the track reconstruction were made combining the primary vertex with pairs of pixel hits. Then, tracks having $p_T > 1\,\text{GeV}$ were extended in the outer tracker using a Kalman filter. Only tracks close to the leading jets were reconstructed.

In Run 2, an iterative tracking is used. In each iteration the following steps are performed:



(A) MHT resolution.



(B) HT resolution.

FIGURE 3.13: MHT (A) and HT (B) resolution, defined as the ratio of the difference of the reconstructed and generated values to the generated one. The plots have been produced using all calorimetric jets above 20 GeV (blue) or only jets passing the fast pile-up identification (red). The plots show that the fast pile-up jet identification improves the resolution of 25 – 30 % and largely reduces the bias. The HT plot has been produced with VBF H→ bb simulation and the MHT plot using $Z(\nu\nu)H(bb)$ simulation with $\sqrt{s} = 13$ TeV and pile-up 40. The distributions have been fitted with Gaussian functions whose parameters are shown in the plots.

1. triplets of pixel hits, or pairs plus the primary vertex, are identified (seeds);
2. the seeds are extended with a Kalman filter to the outer tracker;
3. good quality tracks are selected;
4. pixel and strip hits that are compatible with the tracks are removed from the collection.

The new tracking uses three iterations:

1. in the first iteration, the pixel tracks already reconstructed in the fast pixel tracking are used as seeds ($p_T > 0.9$ GeV);

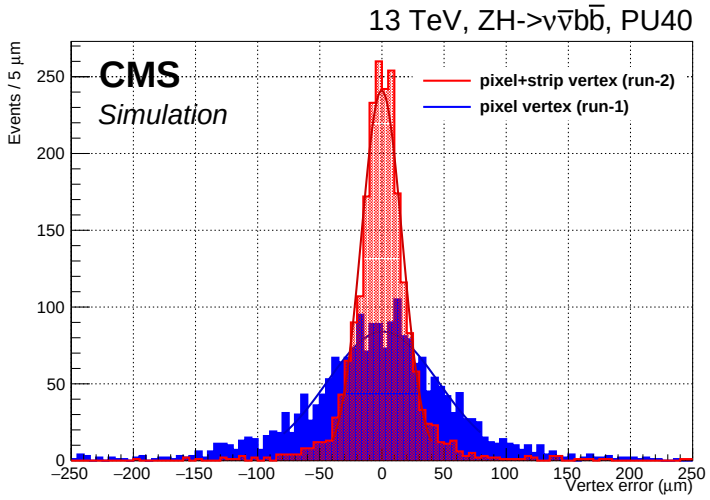


FIGURE 3.14: Comparison of the vertex resolution obtained with the full-tracker tracks (red, Run 2) and pixel tracks (blue, Run 1). The plot has been produced using $t\bar{t}$ simulation with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. The fitted lines correspond to two Gaussian with $\sigma_{\text{old}} = 43\,\mu\text{m}$ and $\sigma_{\text{new}} = 16\,\mu\text{m}$.

2. in the second iteration, additional triplets of pixel hits with $p_T > 0.5\,\text{GeV}$ are used as seeds;

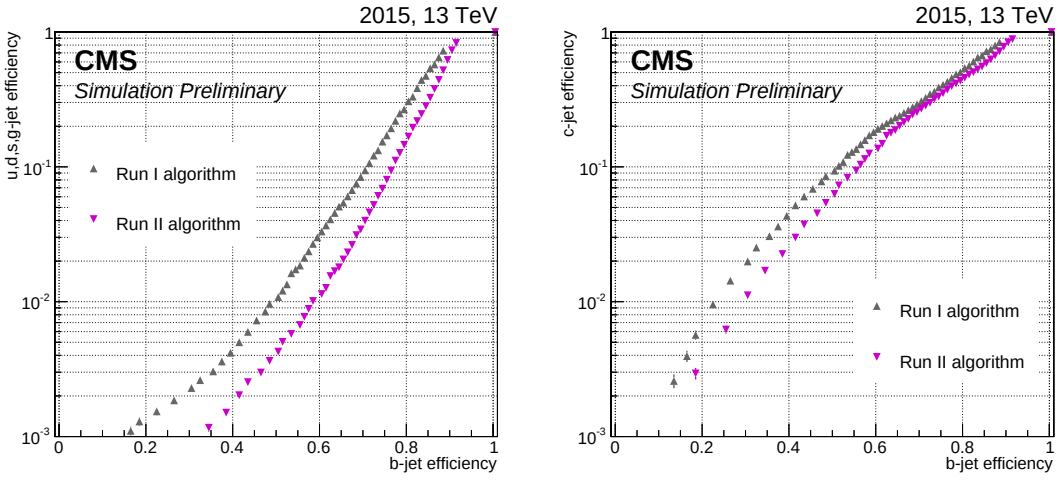
3. in the last iteration, pairs of pixel hits having $p_T > 1.2\,\text{GeV}$ compatible with the primary vertex are used as seeds.

At each iteration, only tracks near the primary vertex are reconstructed. In Run 2, these tracks are also used to fit again the primary vertex obtaining a resolution of about $15\,\mu\text{m}$. Figure 3.14 compares the vertex resolution obtained with the full-tracker tracks (Run 2) and pixel tracks (Run 1).

Eventually, vertex and tracks are used to evaluate the b tagging discriminant. In Run 2 a new offline b-tagging algorithm is used, described in Section 2.3.8, and it has been ported to HLT. Figure 3.15 compares the online b-tagging performance during Run 1 and Run 2: for a b-tagging efficiency of $\sim 60\%$, a typical working point, in Run 2 the mistag rate has been decreased by about a factor three. This improvement is the combined effect of the new vertexing, the new iterative tracking, and the new b-tagging algorithm. In addition, the time needed to run the online b tagging has been reduced of 20% with respect to Run 1, as shown in Figure 3.16.

3.4.1 B-tag performance using 13 TeV data

The new online b tagging has been commissioned with the first 13 TeV data and the performance has been measured directly from data. Figure 3.17a shows the offline b-tagging discriminator distribution for all and online b-tagged jets. The online b-tagging



(A) Efficiency vs fake-rate on u, d, s, and gluon-jets.

(B) Efficiency vs fake-rate on c-jets.

FIGURE 3.15: The plot shows the performance of the Run-1 and Run-2 online b tagging. The improvement is the combined effect of the new b-tag algorithm, the new vertexing, and the new iterative tracking. For a typical working point (b efficiency $\sim 60\%$) the mistag rate on light jets has been reduced by almost a factor three. The plot has been obtained using 13-TeV $t\bar{t}$ simulation with pile-up 40 and bunch spacing 25 ns.

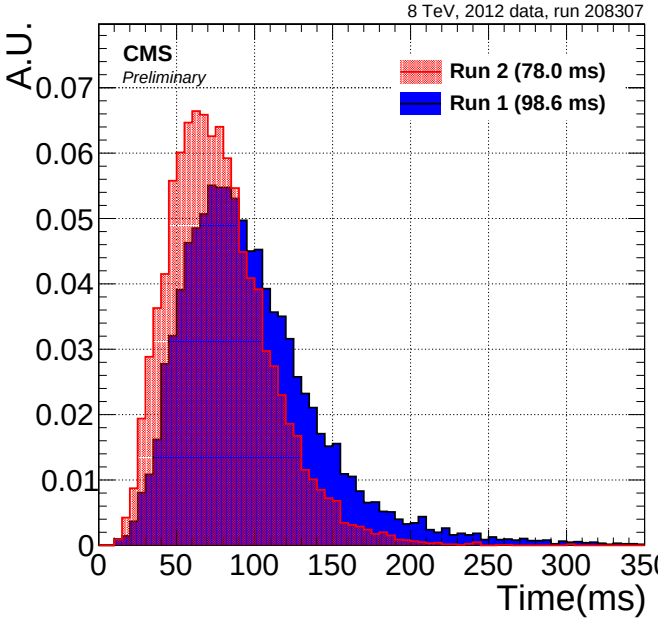
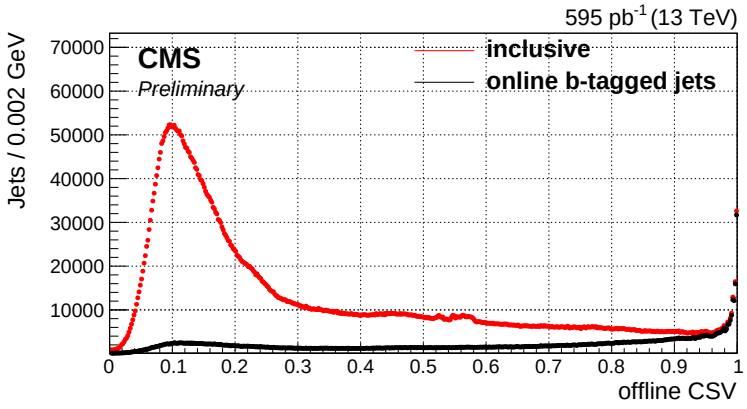
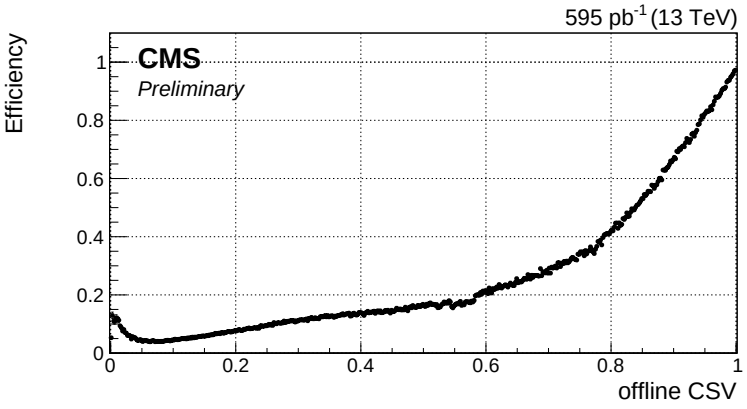


FIGURE 3.16: Distribution of the total time needed to run the improved (red) and default (blue) online b tagging. The time is largely due the tracking in the outer tracker. The measurement was performed using 8 TeV data, selecting events with at least two central calorimetric jets with $p_T > 40$ GeV.



(A) Distribution of the offline b-tagging discriminant for all (red) and online b-tagged (black) jets.



(B) Online b-tagging efficiency as a function of offline b-tagging discriminant.

FIGURE 3.17: The online b-tagging performance using Run-2 data. Figure (A) is the offline b-tagging discriminator distribution for all jets (red) and online b-tagged jets (black). Figure (B) is the ratio between the two distributions. As expected, the online b-tagging efficiency is high for jets with high b-tagging discriminant and low for the other jets. These plots have been obtained with the events collected by a hadronic trigger that required $\sum_i p_i^T > 800\,\text{GeV}$. The online b tagging has been evaluated on the eight leading jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$.

efficiency as a function of the offline b-tagging discriminant is shown in Figure 3.17b. As expected, the online b-tagging efficiency is high for jets with high b-tagging discriminant and low for the other jets.

While in this thesis the main focus is on $ZH \to \nu\nu bb$, I used the improved online b tagging also to design triggers dedicated to the search for the $H \to bb$ produced through the vector boson fusion process [74], and the search for a resonance decaying to $HH \to 4b$ [73], enlarging significantly the trigger acceptance of these searches.

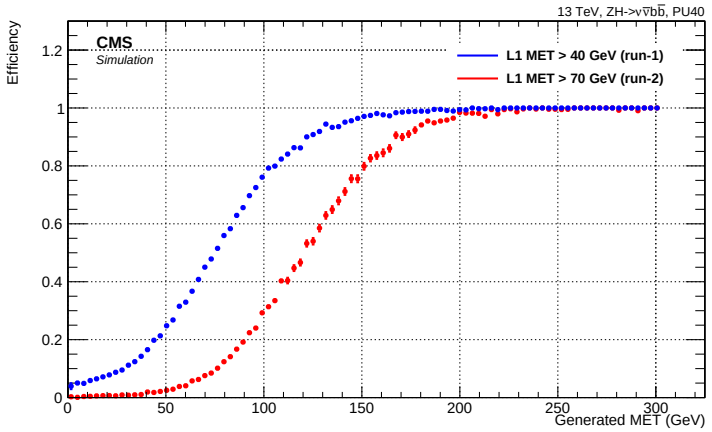


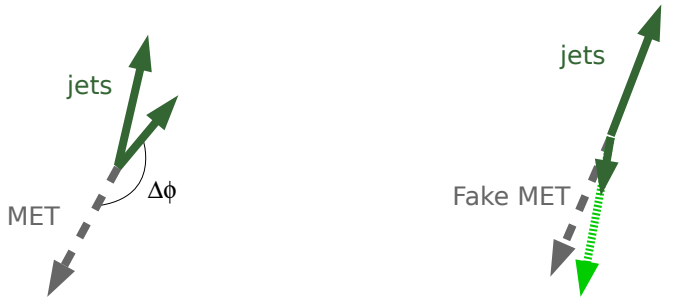
FIGURE 3.18: L1 MET trigger efficiency as a function of the generated MET, using the Run 1 (blue) and high luminosity Run-2 (red) thresholds. The plot has been produced using Z($\nu\nu$)H(bb) simulation with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40.

3.5 Improvements to the online MET reconstruction

In addition to b tagging, the main component of the Z($\nu\nu$)H(bb) triggers is the MET selection. Since MET is a global variable, it is affected by the increase of total energy measured in a single bunch crossing and hence the acceptance of MET triggers is significantly reduced during Run 2, because of the higher pile-up, energy, and luminosity. This section presents some improvements that I have obtained in the reconstruction of the MET in each stage of HLT in order to enlarge the acceptance of the Z($\nu\nu$)H(bb) analysis.

3.5.1 L1 MET

Every HLT MET path is seeded by the L1 MET trigger. In order to keep the rate around 5–10 kHz, the L1 MET threshold has to be increased dramatically during the high luminosity Run 2 ($\mathcal{L} = 1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, PU=40), with respect to Run 1. Figure 3.18 compares the L1 trigger efficiency, as a function of the generated MET, for the nominal cut L1 MET $> 40\,\text{GeV}$ and L1 MET $> 70\,\text{GeV}$. These two thresholds correspond to, respectively, the threshold used during the Run 1 and the one expected to obtain the same rate in the high luminosity Run-2 scenario. In the latter trigger, the L1 MET trigger efficiency is larger than 50% only for offline MET $\gtrsim$ 120 GeV and above 90% for offline MET $\gtrsim$ 170 GeV. As a reference, the search for Z($\nu\nu$)H(bb) in Run 1 was performed starting from MET $> 100\,\text{GeV}$ [27]. The sensitivity of this analysis would be severely limited in Run 2 by this L1 trigger efficiency. As in the L1 trigger tracks are not available, we cannot improve the L1 MET resolution using the tools developed, such as the fast pile-up identification described in Sect. 3.3. Therefore, we need new ideas to improve the efficiency of the L1 MET trigger.



(A) Real MET. (B) Fake MET.

FIGURE 3.19: Diagram (A) represents a generic event having real MET: jets and MET are roughly back-to-back on the transverse plane and the $\Delta\varphi$ angle is around π. Diagram (B) shows a di-jet event with fake MET: here $\Delta\varphi$ angle between jet and MET is near zero because the fake-MET is produced close to the mismeasured jet.

The main contribution to the L1 MET trigger rate is QCD multijet production, when the energy of a jet is mismeasured, faking large MET values. The sketches in Figure 3.19 compare the topologies of events of Z($\nu\nu$)H(bb) and multijet: in the first case, all jets are typically to be back-to-back to the MET on the transverse plane; in multijet events, the MET is expected to be close to the mismeasured jet. Therefore, the $\Delta\varphi$ angle between jets and MET can be used to discriminate these events. Figure 3.20 shows the distribution of this angle reconstructed using the L1 objects, for events having real MET (Z($\nu\nu$)H(bb)) and fake MET (multijet).

In the high luminosity Run 2, the L1 fake MET can be suppressed rejecting events having, at least, one jet with $p_T > 50$GeV and $\Delta\varphi(\text{jet}, \text{MET}) < 1$. This cut has an efficiency of about 80% on signal while reducing the rate of about a factor three. The L1 MET trigger had the minimal unprescaled threshold of L1 MET > 70 GeV. The improved one uses L1 MET > 65 GeV, with the fake-MET veto, plus L1 MET > 80 GeV, without the veto. The two configurations have similar trigger rates and their efficiencies have been compared in Figure 3.21. The improved trigger is more efficient and, roughly, its efficiency curve has been shifted of 5 GeV.

The dependence of the MET trigger rate and efficiency on pile-up is shown in Figure 3.22. The efficiency for Z($\nu\nu$)H(bb) events with generated MET > 120 GeV stays constant, in spite of the veto, because most of the pile-up jets have $p_T \lesssim 30$ GeV.

In 2015, LHC did not reach the expected high luminosity ($1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$) but only $\mathcal{L} = 5 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$, allowing to decrease the L1 MET trigger threshold down to L1 MET > 50 GeV without any veto. The problem of the high L1 MET threshold appeared in 2016, when the luminosity exceeded $1.0 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$. The strategy of the $\Delta\varphi(\text{jet}, \text{MET})$ veto to decrease rate and MET threshold is still valid and has been proposed again.

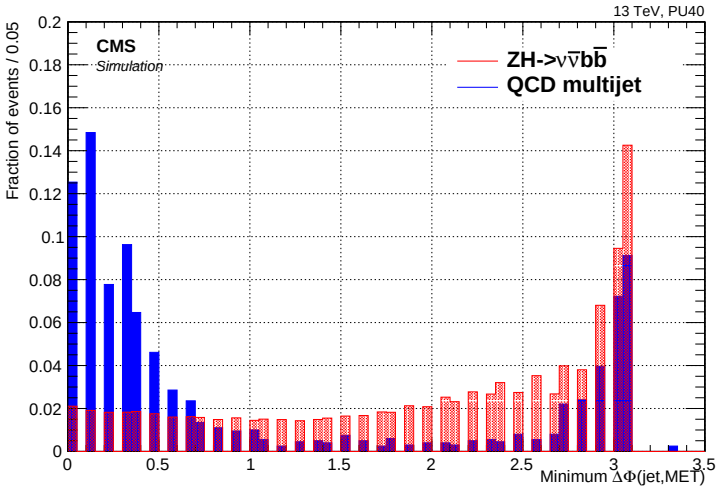


FIGURE 3.20: Distribution of the minimum $\Delta\varphi$ angle between jet and MET for Z($\nu\nu$)H(bb) signal (red) and QCD multijet (blue), normalized to 1. Both samples have been generated with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. The plot has been produced selecting only jets with $p_T > 50\,\text{GeV}$. The variables have been reconstructed using a simulation of the L1 trigger. The bin with $\Delta\varphi = 3.4$ contains events without jets above the threshold. As expected, events with real MET tend to have large $\Delta\varphi(\text{jet}, \text{MET})$, whereas fake-MET events have the opposite behavior. However, a fraction of multijet events has large $\Delta\varphi(\text{jet}, \text{MET})$ as well because the mismeasured jet is too soft. The multijet events are largely dominating in data.

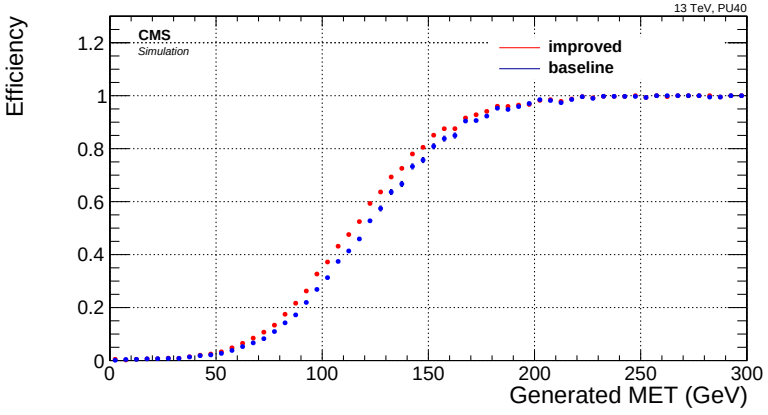
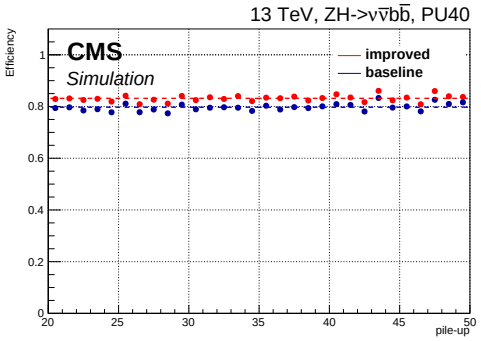
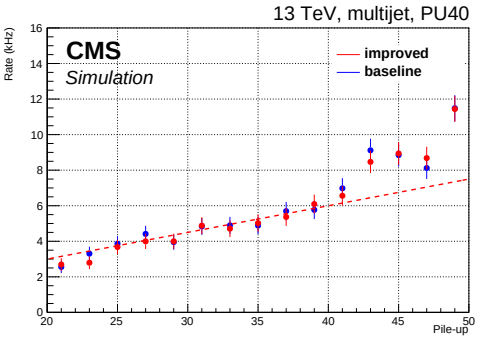


FIGURE 3.21: L1 MET trigger efficiency as a function of the generated MET of Z($\nu\nu$)H(bb) sample with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. In blue it is reported the efficiency of the default trigger ($\text{L1 MET} > 70\,\text{GeV}$). In red it is shown the efficiency of the improved trigger ($\text{L1 MET} > 65\,\text{GeV}$ with the $\Delta\varphi(\text{jet}, \text{MET})$ veto plus L1 MET $>$ 80 GeV).



(A) Efficiency as a function of the pile-up.



(B) Rate as a function of pile-up.

FIGURE 3.22: Efficiency (A) and rate (B) as a function of the number of pile-up vertices for Z($\nu\nu$)H(bb), with a preselection MET $> 120\,\text{GeV}$ on the generated variable, and QCD multijet events. Both samples have been generated with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. The plots have been produced for the improved (red) and for baseline (blue) L1 MET trigger. Figure (A) shows that efficiency is independent on pile-up, even using the fake-MET veto. Figure (B) plots the trigger rates as a function of the pile-up and a linear extrapolation. Both trigger rates increase linearly except in the very high pile-up region. Here, trigger rates increase more than linearly due to pile-up effects.

3.5.2 Calorimetric MET

The final states of $Z(\nu\nu)H(bb)$ is characterized by the presence of jets and high MET. Given the topology of the signal, we expect to have also large MHT, that is the MET evaluated on jets ($\overrightarrow{\text{MHT}} = -\sum_i \text{jet}\ \vec{p}_T^{\,i}$). At HLT, one of the main differences between the calorimetric MET and MHT is that, as the MHT is based on jets, it is more robust against pile-up and can be improved by pile-up jet cleaning. In addition, the MHT uses the jet corrections that improve the MHT resolution considerably. As discussed in Section 3.3.1, the pile-up cleaning can be obtained exploiting the fast pile-up identification and this improves the calorimetric MHT resolution of 25 – 30 %.

Figure 3.23 is a comparison of the resolution of the calorimetric MET, MHT, and MHT with the fast pile-up cleaning, and the L1 MET. As expected, the L1 MET has the worst performance, and the MHT with pile-up cleaning has the best resolution.

3.5.3 PF MET

The PF is the most precise, and slow, reconstruction available at HLT. The PF MET performance, indeed, is better than the calorimetric MET. Likewise to the calorimetric MHT case (Sect. 3.5.2), we exploit the PF MHT in the attempt to further improve the performance. In this case, we can clean the PF jets from noise using the following requirements (jet identification):

1. Neutral hadron energy fraction $< 90\%$;

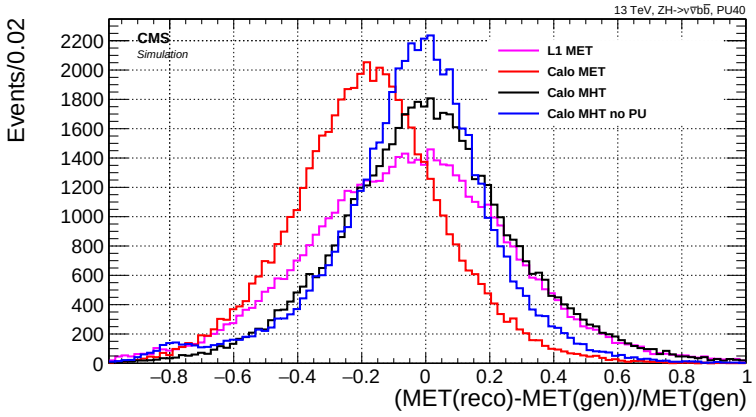


FIGURE 3.23: Resolution of the L1 MET (magenta) and the HLT calorimetric MET (red), MHT (black) and MHT after the fast pile-up cleaning (blue). The resolution has been defined as the difference of the reconstructed and generated values to the generated one. The plots have been obtained using $Z(\nu\nu)H(bb)$ simulation with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. The MHT evaluated on jets after the fast pile-up cleaning has the best resolution. The L1 MET values have been multiplied by a factor 1.5 because they systematically underestimate the true MET value.

2. Electron energy fraction $< 99\%$;
3. Photon energy fraction $< 99\%$;
4. Number of constituent ≥ 2;
5. Number of charged constituent ≥ 1;

where the fractions refer to the PF jet energy components (Sect. 2.3.7). The charged constituents are originating from the primary vertex. The jet identification cuts are similar to what we will require in the offline $Z(\nu\nu)H(bb)$ analysis (Sect. 4.7) to distinguish jets from noise, e.g. spikes in the calorimeter, or other objects, e.g. electrons.

The resolutions of MET, MHT after the jets selection, and MHT are shown in Figure 3.24. As expected, the MHT with the jet identification is the variable with the best resolution.

3.6 Triggers for $H(bb)$ in Run 2

3.6.1 Triggers for the search for $Z(\nu\nu)H(bb)$

At the beginning of the preparation of the HLT menu for the high luminosity Run 2, the only trigger available for the $Z(\nu\nu)H(bb)$ channel was the pure MET trigger `HLT PFMET170`. This trigger was obtained increasing the MET threshold with respect to the Run-1 trigger, in order to obtain a sustainable trigger rate in high luminosity scenario. The `HLT PFMET170` uses the following cuts:

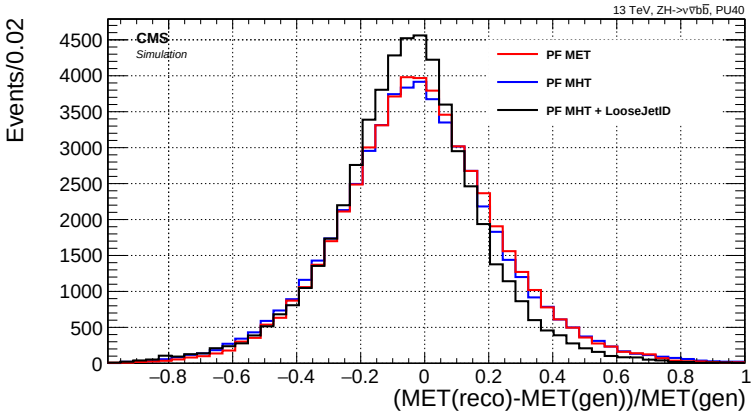


FIGURE 3.24: Resolution of the online PF MET (red), PF MHT (blue) and PF MHT after the jet identification (black). The resolution has been defined as the ratio of the difference of the reconstructed and generated values to the generated one. The plots have been obtained using $Z(\nu\nu)H(bb)$ simulation with $\sqrt{s} = 13\,\text{TeV}$ and pile-up 40. The PF MHT evaluated on jets after the jet identification has the best resolution.

- L1 MET $> 70\,\text{GeV}$;
- calorimetric MET $> 90\,\text{GeV}$;
- PF MET $> 170\,\text{GeV}$.

This trigger is expected to have a rate of $\sim 10\,\text{Hz}$ in the high luminosity scenario ($\mathcal{L} = 1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, PU=40).

This and the following trigger rates have been estimated with simulations using a method that I developed. This method solved the problem of the rate double-counting due to pile-up collision in simulations and is described in Appendix B.

In order to enlarge the trigger acceptance of the $Z(\nu\nu)H(bb)$ analysis, a new trigger has been developed exploiting the new PF MHT variable, using the jet identification, and the new L1 MET seed, with the $\Delta\varphi(\text{jet}, \text{MET})$ veto to reject the QCD multijet events. As a $Z(\nu\nu)H(bb)$ event has typically both large MET and MHT, the new trigger cuts on both variables with the same threshold. The `HLT PFMET120 PFMHT120 IDTight` is based on the following filters:

- L1 MET $> 70\,\text{GeV}$ or L1 MET with QCD veto $> 60\,\text{GeV}$;
- calorimetric MET $> 90\,\text{GeV}$ and calorimetric MHT $> 90\,\text{GeV}$;
- PF MET $> 120\,\text{GeV}$ and PF MHT using jet identification $> 120\,\text{GeV}$.

The expected trigger rate in the high luminosity scenario is $\sim 28\,\text{Hz}$.

Finally, a trigger with online b tagging for the search of $Z(\nu\nu)H(bb)$ has been added to the HLT menu in order to further increase the trigger acceptance at low MET. The trigger, `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72`, exploits also the calorimetric MHT using the fast pile-up identification. It is made up from:

- L1 MET $> 70\,\text{GeV}$ or L1 MET with QCD veto $> 60\,\text{GeV}$;
- calorimetric MHT $> 90\,\text{GeV}$ with the fast pile-up cleaning;
- one online b jet;
- PF MET $> 90\,\text{GeV}$ and PF MHT using jet identification $> 90\,\text{GeV}$.

The expected rate for this trigger is about $\sim 28\,\text{Hz}$. A prescaled control trigger without b tagging has been added to the HLT menu as well.

In 2015, the LHC never reached the expected high luminosity. For this reason, it has been possible to reduce the trigger thresholds and, in particular, the trigger `HLT PFMET90 PFMHT90 IDTight` has been added to the HLT menu:

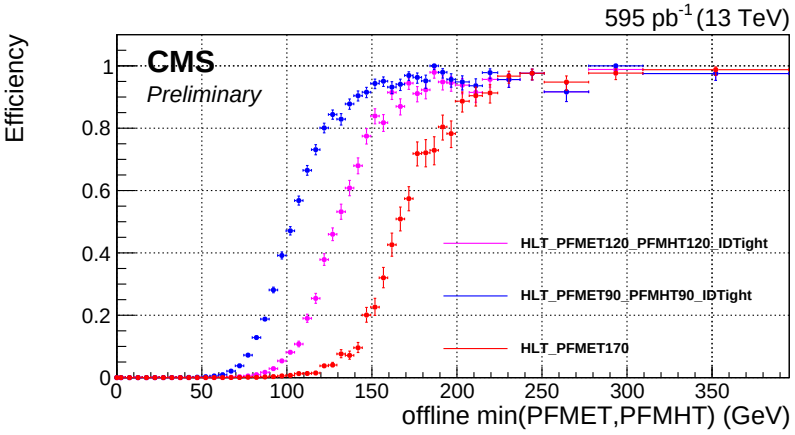
- L1 MET $> 50\,\text{GeV}$;
- calorimetric MET $> 70\,\text{GeV}$ and calorimetric MHT $> 70\,\text{GeV}$;
- PF MET $> 90\,\text{GeV}$ and PF MHT with jet identification $> 90\,\text{GeV}$.

Figure 3.25 is a comparison of the trigger efficiencies measured using the data collected by a single electron trigger in 2015. The Figure shows that `HLT PFMET120 PFMHT120 IDTight` acceptance has been greatly increased compared to `HLT PFMET170`. The bottom plot shows that the trigger acceptance for the $Z(\nu\nu)H(bb)$ channel in the high luminosity scenario will be almost unchanged, thanks to the `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72` trigger.

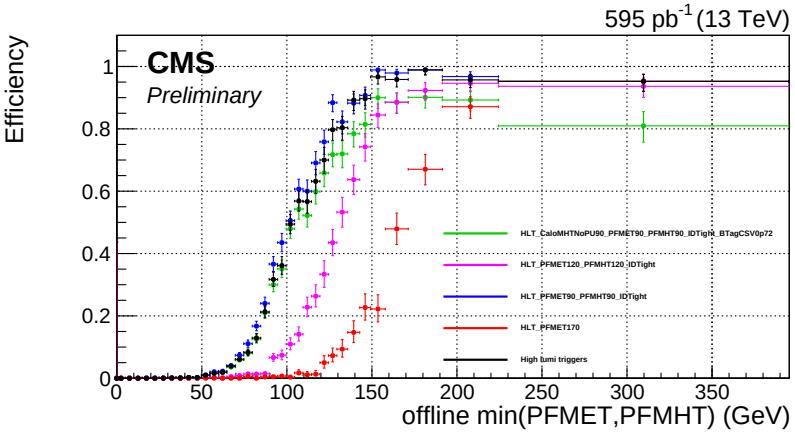
The trigger rates measured at the luminosity peak of 2015 ($\mathcal{L} = 5.2 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$, PU$\sim$ 17) were:

- `HLT PFMET170`: $\sim 14\,\text{Hz}$;
- `HLT PFMET120 PFMHT120 IDTight`: $\sim 4\,\text{Hz}$;
- `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72`: $\sim 2.2\,\text{Hz}$;
- `HLT PFMET90 PFMHT90 IDTight`: $\sim 14\,\text{Hz}$;
- `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight`: $\sim 0.7\,\text{Hz}$ (prescale factor was 17);

Figure 3.26 shows the rate of `HLT PFMET90 PFMHT90 IDTight` as a function of time during a run of November 2^{nd}, 2015.



(A) Single electron dataset.



(B) Single electron dataset with one offline b-tagging selection.

FIGURE 3.25: The efficiency of MET triggers as a function of the minimum between MET and MHT. Both plots have been obtained using a single electron trigger with the preselection: two offline jets with $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$, one offline electron with $p_T > 20\,\text{GeV}$. In Figure (B) one tight offline b tagging has been required too. "High lumi triggers" means `HLT PFMET90 PFMHT90 IDTight OR HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72`.

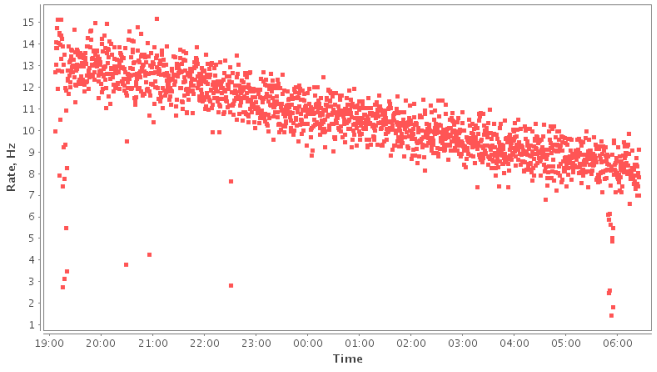


FIGURE 3.26: The rate of HLT PFMET90 PFMHT90 IDTight as a function of time in a run of November 2nd, 2015. At the beginning of the run, the instantaneous luminosity was $\sim 5.2 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ and the pile-up ~ 17. At the end of the run, the instantaneous luminosity was $\sim 4.0 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$ and the pile-up ~ 13.

3.6.2 Triggers for the search for $X \to H(bb)H(bb)$

One of the Run 2 analyses which have most benefited from the improved online b tagging is the search for a resonance decaying to $HH \to 4b$. Two triggers have been developed for this channel and they select events requiring, at least, four jets and three online b-tagging in the event.

The first trigger, `HLT QuadJet45 TripleBTagCSV0p67`, requires the following cuts:

- two L1 jets with $p_T > 84\,\text{GeV}$ and $|\eta| < 3.0$;
- two central[1] calorimetric jets with $p_T > 90\,\text{GeV}$ and two with $p_T > 30\,\text{GeV}$;
- three b tagged jets with CSV > 0.67;
- two central PF jets with $p_T > 90\,\text{GeV}$ and two with $p_T > 30\,\text{GeV}$.

The second trigger, `HLT QuadJet45 TripleBTagCSV0p67`, requires the following cuts:

- four L1 jets with $p_T > 40\,\text{GeV}$ and $|\eta| < 3.0$;
- four central calorimetric jets with $p_T > 45\,\text{GeV}$;
- three b tagged jets with CSV > 0.67;
- four central PF jets with $p_T > 45\,\text{GeV}$.

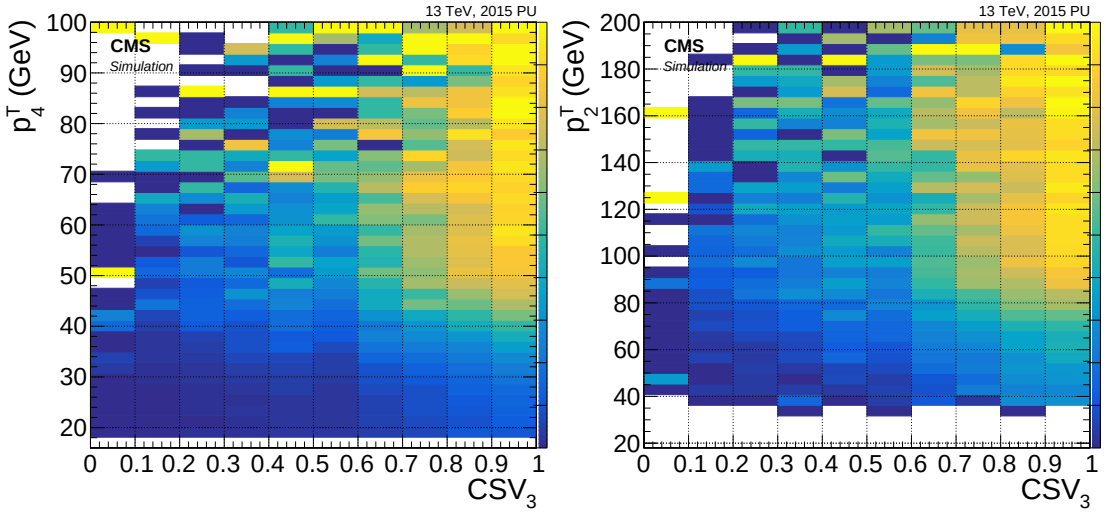
Both triggers are expected to be unprescaled and to have a rate of $\sim 10\,\text{Hz}$ each in high luminosity scenario ($\mathcal{L} = 1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$, PU=40). The performance of the two triggers is reported in Fig. 3.27.

The triggers were online during the 2015 data-taking. The data collected have been used to search for a resonance $HH \to 4b$ and the results are reported in [73]. Figure 3.28 shows the trigger efficiency as a function of the resonance mass, after the offline selection used in that analysis.

3.6.3 Triggers for the search of $H(bb)$ produced through vector-boson fusion

We developed a trigger with b tagging dedicated to the search for the $H \to bb$ produced through vector-boson fusion (VBF). The topology of this signal is composed of two b jets, originating from the Higgs boson decay, and other two jets with large $\Delta\eta$ between them, produced by the quarks that scattered off the vector bosons. The trigger, `HLT`

[1] with $|\eta| < 2.4$



(A) (B)

FIGURE 3.27: Trigger efficiency of `HLT QuadJet45 TripleBTagCSV0p67` as a function of the fourth leading offline central jet p_T and of the third leading offline jet CSV (A). Trigger efficiency of `HLT DoubleJet90 Double30 TripleBTagCSV0p67` as a function of the second leading offline central jet p_T and of the third leading offline jet CSV (B). The plot have been obtained in using a multijet simulation enriched of b jets. The scale is comprised between 0% (dark blue) and 100% (yellow) and is reported on the right of each plot.

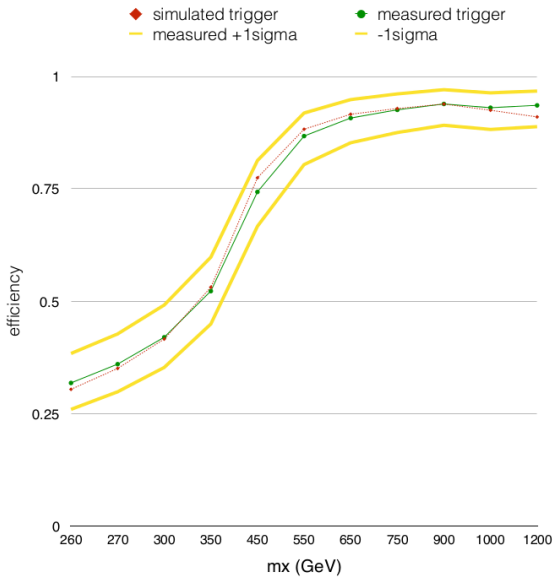


FIGURE 3.28: Trigger efficiency as a function of the resonance mass, after the offline analysis selection [73].

`QuadPFJet DoubleBTagCSV VBF Mqq240`, requires four jets and uses a double b tagging to filter events and to identify the Higgs boson jets. The $\Delta\eta$ and invariant mass (m_{jj}) between the two un-tagged jets are used to select events with the VBF topology. The cuts used are the following:

- three L1 jets with $p_T^1 > 92\,\text{GeV}$, $p_T^2 > 76\,\text{GeV}$, and $p_T^3 > 64\,\text{GeV}$;
- four calorimetric jets with $p_T^1 > 80\,\text{GeV}$, $p_T^2 > 65\,\text{GeV}$, $p_T^3 > 50\,\text{GeV}$, and $p_T^4 > 15\,\text{GeV}$, with two of them having $|\Delta\eta| > 1.5$ and $M_{qq} > 150\,\text{GeV}$.
- one jet with CSV > 0.74, performing b tagging with the regional tracking;
- four PF calorimetric jets with $p_T^1 > 92\,\text{GeV}$, $p_T^2 > 76\,\text{GeV}$, $p_T^3 > 64\,\text{GeV}$, and $p_T^4 > 15\,\text{GeV}$; two of them are b-tagged using the PF tracks ($\text{CSV}_1 > 0.68$, $\text{CSV}_2 > 0.58$) and are required to have $|\Delta\varphi| < 2.4$ in order to reduce the di-b-jet background; the remaining two jets are required to have $|\Delta\eta| > 2.0$ and $m_{jj} > 240\,\text{GeV}$.

In order to recover the small online b tagging inefficiency, another trigger with a similar strategy has been prepared requiring only one online b tagging but a larger m_{jj} cut (`HLT QuadPFJet SingleBTagCSV VBF Mqq500`). These trigger are expected to have a rate of $\sim$ 15 Hz each at $1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$ of instantaneous luminosity. Since the peak luminosity of 2015 was only $5 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$, we decreased the trigger cut thresholds deploying the triggers `HLT QuadPFJet DoubleBTagCSV VBF Mqq200` and `HLT QuadPFJet SingleBTagCSV VBF Mqq460`. Figure 3.29 shows the efficiency of `HLT QuadPFJet DoubleBTagCSV VBF Mqq200` trigger as a function of the offline $\Delta\eta$ and m_{jj} of the two un-tagged jets.

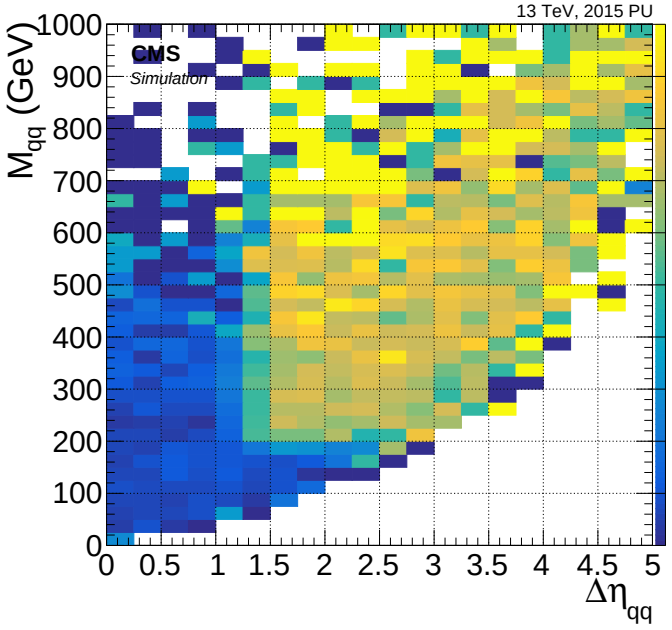


FIGURE 3.29: Trigger efficiency of `HLT QuadPFJet DoubleBTagCSV VBF Mqq240` as a function of the offline $\Delta\eta$ and m_{jj} of the two offline un-tagged jets. The plot have been obtained in using a multijet simulation enriched of b jets. The scale is comprised between 0% (dark blue) and 100% (yellow) and is reported on the right of the plot.

Chapter 4

The search for $Z(\nu\nu)H(bb)$

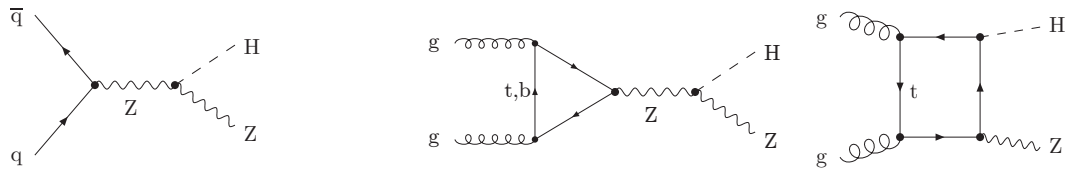
This chapter presents the search for the $Z(\nu\nu)H(bb)$ final state performed using the first 13 TeV *data collected in 2015 by the triggers that I developed during the Long Shutdown 1 (Chap.3). Compared to the Run-1 analysis [27], this analysis presents some new features that will be emphasized, like a new signal extraction fit and a new treatment of the QCD multijet background. The search has been performed using* $2.32\,\text{fb}^{-1}$ *of data collected in 2015 and the expected sensitivity after the 2016 data-taking (*$\sim 20\,\text{fb}^{-1}$*) has been estimated.*

4.1 Introduction

The $H \to bb$ is the Higgs decay mode with the largest branching ratio. It has not been observed yet because of the large QCD multijet background that overwhelms the $H \to bb$ produced through the gluon-gluon and vector-boson fusion. In the other production channels, Z/WH and ttH, the multijet background can be reduced requiring a leptonic decay of the Z/W boson or of the top quark.

This chapter describes the search for the $H \to bb$ in the associated production with a Z boson decaying to neutrinos. This is one of the most effective channels among the searches for the $H \to bb$, because it benefits of both the higher cross section of the ZH production, compared to the ttH channel, and of the large branching ratio of the $Z \to \nu\nu$, with respect to the $Z \to$ charged leptons. The fraction of the $pp \to Z(\nu\nu)H(bb)$ events to the inclusive Higgs production is about 0.20%.

The ZH final state can be produced from the quark-antiquark ($q\bar{q}$) and gluon-gluon via loops (gg) interaction. At $\sqrt{s} = 13\,\text{TeV}$, the expected cross-section times branching ratio for the $q\bar{q} \to ZH \to \nu\nu bb$ is of $(0.088 \pm 0.004)\,\text{pb}$ and for the gluon-gluon fusion initiated process is of $(0.014 \pm 0.004)\,\text{pb}$. The leading order Feynman diagrams of the



(A) $q\bar{q}ZH$ (B) $ggZH$

FIGURE 4.1: Leading order diagrams of the Higgs boson production in association with a Z boson, initiated by quark-antiquark (A) and gluon-gluon (B) fusions.

two processes are shown in Fig. 4.1. The interference of the two $ggZH$ diagrams in Fig. 4.1b is destructive and is dominated by the triangle loop [75].

4.1.1 Z/W boson reconstruction

The search for the $ZH \to \nu\nu bb$ is performed together with the searches for the $WH \to \ell\nu bb$ and $ZH \to \ell\ell bb$, and therefore the signal regions of the three searches must be orthogonal, in order to avoid a double counting of the signal events. The events have been divided in five categories, depending on the reconstructed Z/W boson candidate decay:

- $Z \to \mu\mu$: two isolated muons with $p_T > 20, 15\,\text{GeV}$ and opposite sign;
- $Z \to ee$: two isolated electrons with $p_T > 20, 15\,\text{GeV}$ and opposite sign;
- $W \to \mu\nu$: one isolated muons with $p_T > 25\,\text{GeV}$;
- $W \to e\nu$: one isolated electrons with $p_T > 25\,\text{GeV}$;
- $Z \to \nu\nu$: $\text{MET}, \text{MHT} > 120\,\text{GeV}$.

the above conditions are evaluated in the given order and the event is assigned with the first satisfied criterium.

4.2 Signal and background

4.2.1 Signal topology

The $ZH \to \nu\nu bb$ topology is characterized by the presence of two b jets, with an invariant mass close to $m_{jj} \sim 125\,\text{GeV}$, and large MET. The di-jet system and the MET are typically back-to-back and balanced on the transverse plane. Additional jets may arise from the initial state (ISR) or final state (FSR) radiation. An important difference between them is that only FSR jets have to be taken into account in the Higgs mass

reconstruction. Pile-up collisions create soft jets, but they can be tagged and cleaned away effectively. No isolated leptons are expected. Possible leptons may be found close to the b-jet directions, as products of the B hadron decays.

The main difference in signature between the $gg \to ZH$ and $q\bar{q} \to ZH$ productions is that the first one contains more often ISR jets, because of the gluons in the initial state and of the colored particles in the loop (see Fig.4.1b). The $WH \to \ell\nu bb$ signature is identical to the $ZH \to \nu\nu bb$, if the lepton from the W decay is outside the acceptance of the detector or is not reconstructed, and hence is included as signal cross section. The overall signal is composed of $\sim 50\%$ $q\bar{q}ZH$, $\sim 30\%$ $ggZH$, and $\sim 20\%$ WH. These fractions have been evaluated using the signal region defined later in Sect. 4.7.1.

Figure 4.2 shows a signal candidate event recorded in 2015. The event contains two b jets with high b-tagging discriminators, having an invariant mass compatible with the Higgs boson mass. The system is balanced on the transverse plane by large MET. No additional leptons or jets were observed.

4.2.2 Z/W + jets

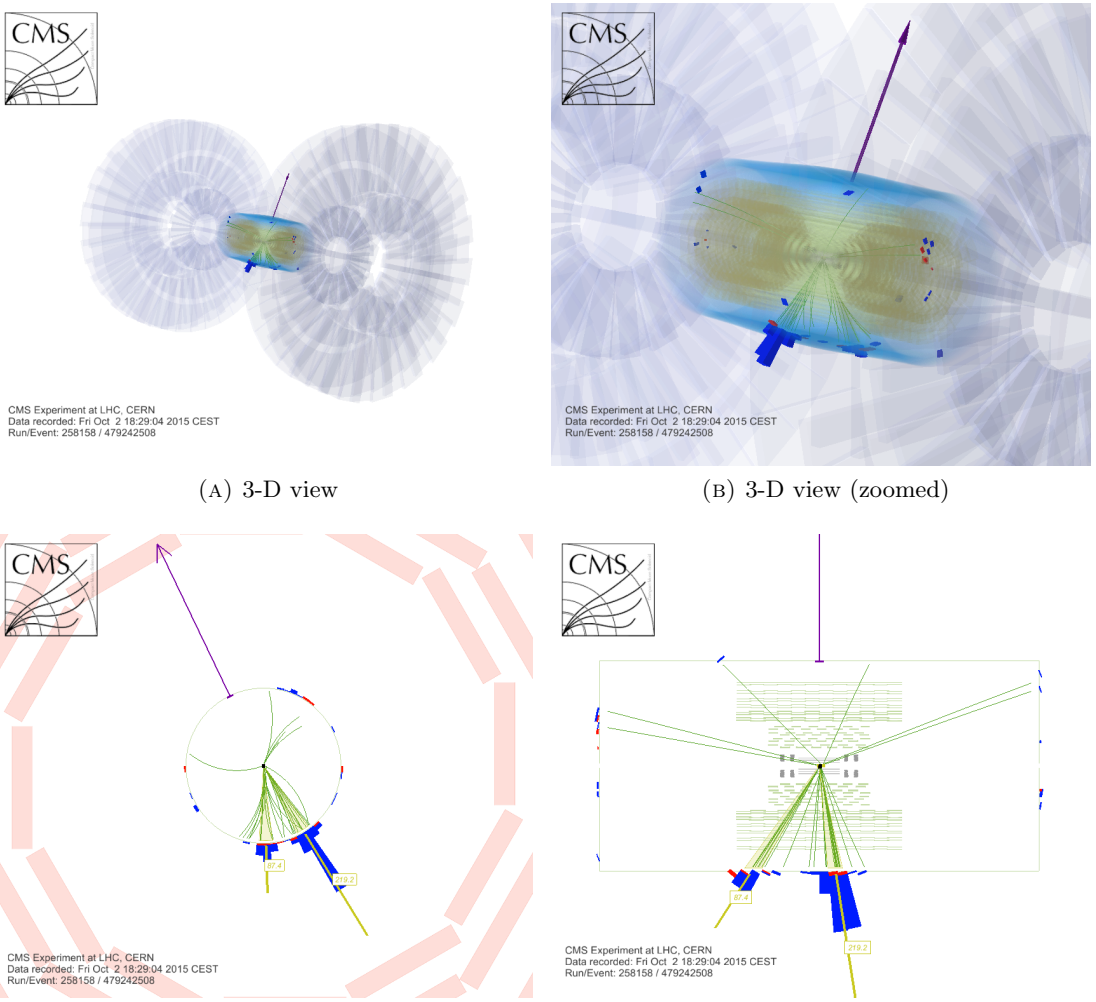
The $Z(\nu\nu)$+jets and the $W(\ell\nu)$+jets productions are the main backgrounds of this analysis. In particular, the $Z(\nu\nu)$+b-jets component has a signature that is identical to the signal one with the exception of the di-jet invariant mass. The remaining components, $Z(\nu\nu)$+light-jets and $W(\ell\nu)$+jets, can be reduced applying, respectively, the b tagging and the lepton veto.

In this analysis, the V+jets (i.e. Z+jets and W+jets) have been split in three components: V+0b−jets, V+1b−jets, and V+2b−jets. The b jets are counted as the number of simulated jets containing a B hadron and having $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$.

The cross sections of the $Z(\nu\nu)$+jets and the $W(\ell\nu)$+jets events with the sum of the parton transverse momenta genHT $> 100\,\text{GeV}$ at $\sqrt{s} = 13\,\text{TeV}$ are, respectively, $\sigma_{Z+\text{jets}} = 458.6\,\text{pb}$ and $\sigma_{W+\text{jets}} = 1772.4\,\text{pb}$.

4.2.3 Top

The $t\bar{t}$ production is the other large background of the $ZH \to \nu\nu bb$ search. The $t\bar{t}$ events are characterized by two W's and two b jets in the final state. The topology is close to the signal if at least one of the W bosons decays leptonically and the lepton from the W decay is outside the detector acceptance or is not reconstructed. Differently from signal, $t\bar{t}$ events have in general more additional jets and, often, at least one jet with a small $\Delta\varphi(\text{jet}, \text{MET})$. The $t\bar{t}$ cross section is large and the $t\bar{t}$ background is dominated by the



(A) 3-D view

(B) 3-D view (zoomed)

(C) Transverse view.

(D) Longitudinal view.

FIGURE 4.2: Display of an event recorded on October 2nd, 2015. The event contains two jets with $p_{T1} = 219.2\,\text{GeV}$ and $p_{T2} = 87.4\,\text{GeV}$, and large b-tagging discriminants. The di-jet system has an invariant mass of 128.0 GeV, compatible with the Higgs boson mass, and $p_T = 296.4\,\text{GeV}$. The di-jet system is balanced on the transverse plane by 308.9 GeV of missing transverse energy. No additional jets with $p_T > 20\,\text{GeV}$ have been observed and no isolated leptons. In the Figures, the green lines are the tracks with $p_T > 1.5\,\text{GeV}$, the yellow lines are the jets with $p_T > 20\,\text{GeV}$, the violet arrow indicates the MET, the blue and red towers correspond to the energy deposited in the HCAL and ECAL, respectively. The Figures show also part of the CMS detectors: the muon chambers (red), ECAL (cyan), strips (green), and the pixels (grey). This event is compatible with the $ZH \to \nu\nu bb$ topology.

configuration with only one W decaying to leptons (85%). The reminder is given by the configuration with two W decaying to leptons (15%).

The single top production is a small background and is dominated by the associated production with a W boson decaying leptonically.

The cross sections are 831.7 pb for the $t\bar{t}$ production and 71.2 pb for the single-top produced in association with a W boson, with a lepton in final state.

4.2.4 VV

The signature of the double vector bosons production ($ZZ/WZ/WW$) is very similar to the signal, especially in the $ZZ \to \nu\nu bb$ decay. Indeed, the only difference in topology of the $ZZ \to \nu\nu bb$ process, with respect to the signal, is that the m_{jj} distribution peaks at $m_Z \sim 91\,\text{GeV}$, instead of $m_H \sim 125\,\text{GeV}$.

The cross sections are $\sigma_{ZZ} = 16.5\,\text{pb}$, $\sigma_{WZ} = 47.1\,\text{pb}$, and $\sigma_{WW} = 115.0\,\text{pb}$.

4.2.5 Multijet

In multijet events, MET is usually produced by the mis-measurement of the energy of a jet. However, MET may be created by the emission inside a jet of a high-p_T neutrino originating from a hadron decay, as shown in Fig. 4.3. The fraction of these events is predicted with large uncertainty since it depends on production rates of hadrons and branching ratios that are known with limited precision.

It has been found that a fake MET is produced when at least one jet points towards malfunctioning ECAL cells, shown in Fig. 4.4. The dead ECAL cells (top plots of Fig. 4.4) have been identified selecting events with MET $> 90\,\text{GeV}$ and inspecting the η vs φ distribution of the sub-leading jet, when it is close to the MET direction. The hot ECAL cells [1] (bottom plots of Fig. 4.4) have been localized selecting events with MET $> 90\,\text{GeV}$ and inspecting the η vs φ distribution of the leading jet, back-to-back to the MET direction. The positions of the dead and hot ECAL cells did not change during the 2015 data-taking.

In a multijets simulation with reconstructed MET $> 120\,\text{GeV}$, $\sim 50\%$ of events contains a neutrino with $p_T > 50\,\text{GeV}$ and $\sim 40\%$ contains an underestimated jet pointing towards to a dead ECAL cell. The two sets of events are roughly independent and the remaining $\sim 10\%$ of multijet events with large MET are due to other mismeasured jets.

The main feature of the multijet events is the presence of at least one jet close to the MET direction on the transverse plane. For these reason, the main variable used in rejecting

[1] An hot ECAL cell is a channel recording fake high energy deposits due to to malfunctioning of the electronics.

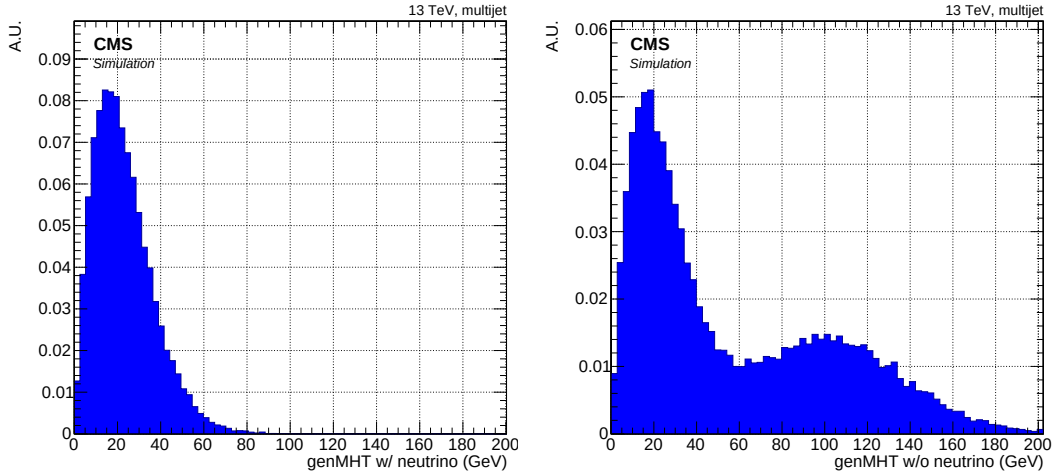


FIGURE 4.3: MET evaluated using generator level information in multijet events with simulated jets with $p_T > 20\,\text{GeV}$, including (left) and excluding (right) the neutrinos originating from the hadron decays in the computation of the MET. The plot has been produced using a multijet simulation with the sum of the generated transverse parton momenta within 500 GeV and 700 GeV and selecting the events with reconstructed MET $> 120\,\text{GeV}$. The plots shows that part of the multijet events with large MET is due to the emission of a high-p_T neutrino, corresponding to the events around $\sim$ 100 GeV on the right plot. In the remaining events, the large MET is produced by the mismeasurement of the transverse momentum of a jet. The peak around 20 Ge is created by the acceptance cut on the simulated jets ($p_T > 20\,\text{GeV}$).

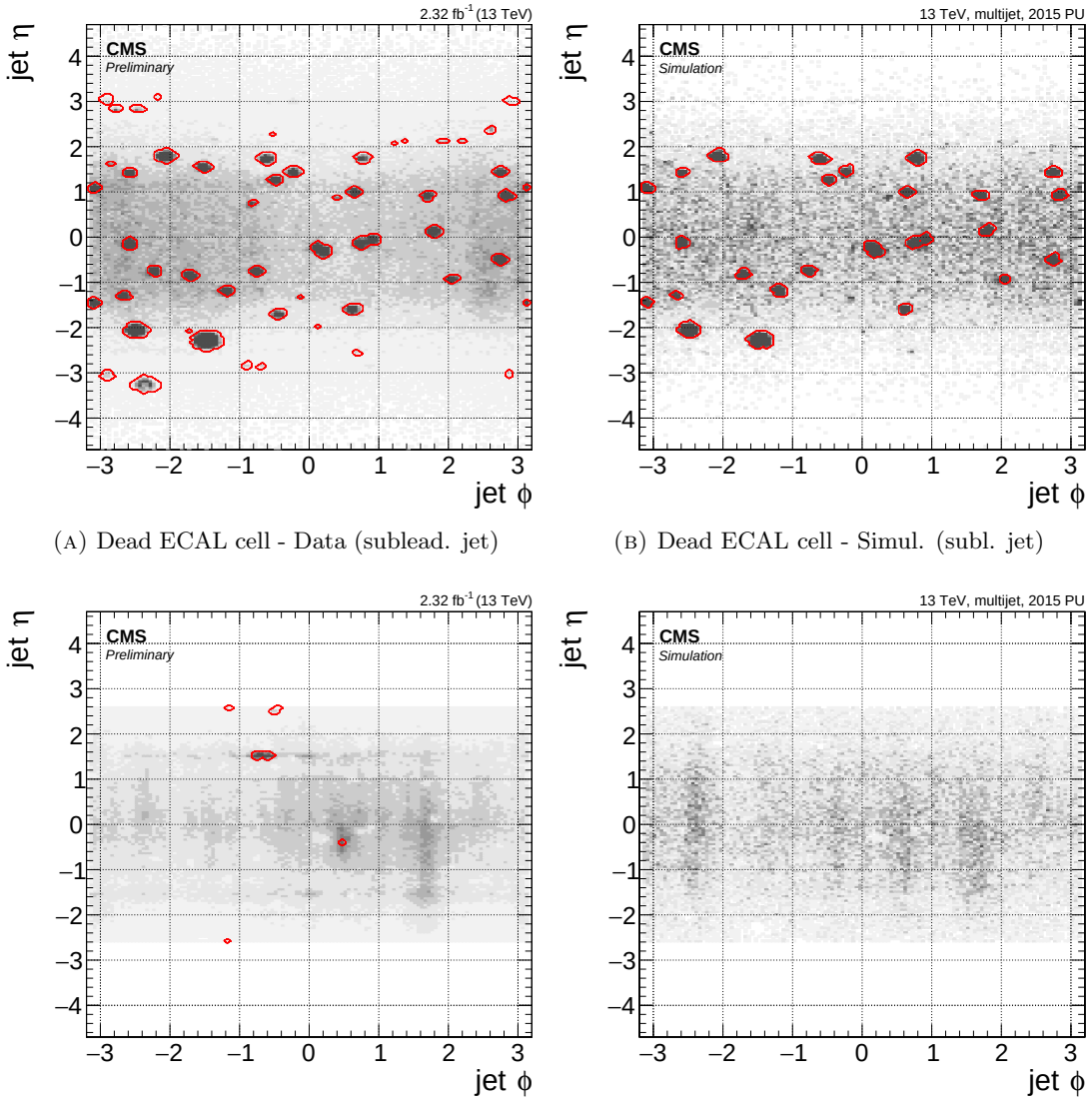
the multijet events is the minimum $\Delta\varphi(\text{jet}, \text{MET})$. Figure 4.5 shows the distribution of $\min(\Delta\varphi(\text{jet}, \text{MET}))$, in multijets and signal events, considering the two leading jets.

The cross section used for multijet events with genHT $> 300\,\text{GeV}$ is $\sigma_{\text{multijets}} = 388.0\,\text{nb}$.

4.3 Analysis strategy

In this analysis, we define a signal region, enriched of signal events, and six control regions, each one enriched of $t\bar{t}$, Z + light-jets, Z + b-jets, W + light-jets, W + b-jets, or QCD multijet. The shapes of the main backgrounds ($t\bar{t}$, W+jets, Z+jets, multijet) are estimated using simulations and their normalizations are fitted from data in the control regions. For the minor backgrounds (single top and $ZZ/ZW/WW$), both the shape and the normalization are taken from the simulations. The control regions are used both to fit the main background normalizations and to test the accuracy of the modeling of kinematic distributions in the simulated samples.

The signal extraction is performed with a simultaneous fit of the distributions of the second highest jet b-tagging discriminant in the control regions and of a multivariate discriminant - trained to separate signal from backgrounds - in the signal regions. In the fit, both the normalization and shape systematics are taken into account using nuisance parameters. In order to increase the sensitivity, the search has been split in two bins



(A) Dead ECAL cell - Data (sublead. jet)

(B) Dead ECAL cell - Simul. (subl. jet)

(C) Hot ECAL cell - Data (leading jet)

(D) Hot ECAL cell - Simulation (lead. jet)

FIGURE 4.4: Distribution of jet η and φ for the sub-leading jet when it is close to the MET direction ($\Delta\varphi(\text{jet}, \text{MET}) < 0.3$) for data (A) and multijet simulation (B). The observed excesses correspond to the dead ECAL cells, and they have been mapped (red line). The plot (A) contains more dead cells than the plot (B) showing that the simulation does not take into account all dead cells. Distribution of jet η and φ for the leading jet when it is back-to-back to the MET direction ($\Delta\varphi(\text{jet}, \text{MET}) > \pi - 0.3$) for data (C) and multijet simulation (D). The observed excesses correspond to the hot ECAL cells, and they have been mapped (red line). Only plot (C) contains hot cells, showing that the hot cells are not reproduced by simulation. The plots have been obtained in samples having MET > 90 GeV.

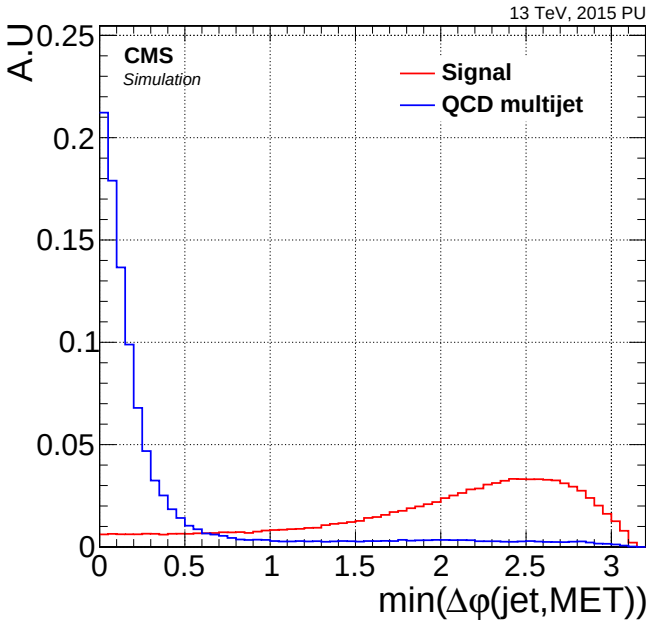


FIGURE 4.5: $\min(\Delta\varphi(\text{jet}, \text{MET}))$ for the two leading jets for signal (red) and multijet (blue) simulation with 500 GeV < genHT < 700 GeV. Events are required to have MET > 120 GeV.

of $p_T(Z)$: 120 GeV $< p_T(Z) <$ 150 GeV and $p_T(Z) >$ 150 GeV. The $p_T(Z)$ has been defined as the minimum between the MET and MHT.

4.4 Data and simulated samples

The search has been performed in a data sample corresponding to 2.32 fb^{-1} of proton-proton collision at a center-of-mass energy of $\sqrt{s} = 13$ TeV. During the data-taking, the number of simultaneous proton-proton collisions was roughly Gaussian-distributed with $\mu \sim 11.4$ and $\sigma \sim 2.1$.

Signal and background events have been simulated using various Monte Carlo (MC) event generators, with the CMS detector response modeled with GEANT4 [76].

The signal processes ($q\bar{q}ZH$,$ggZH$,WH) and the $t\bar{t}$ have been simulated using Powheg [77] and aMC@NLO [78] has been used as an alternative event generator. The Z/W + jets and multijet backgrounds have been simulated with Madgraph [79], and the di-boson ($ZZ/ZW/WW$) with Pythia8 [80]. The single-top background has been produced both with aMC@NLO (s and t channels) and Powheg ($t - W$ channel).

The production cross-sections have been rescaled to the NNLO. The parton hadronization has been performed with Pythia8 and the PDF have been taken from NNPDF3.0 [81]

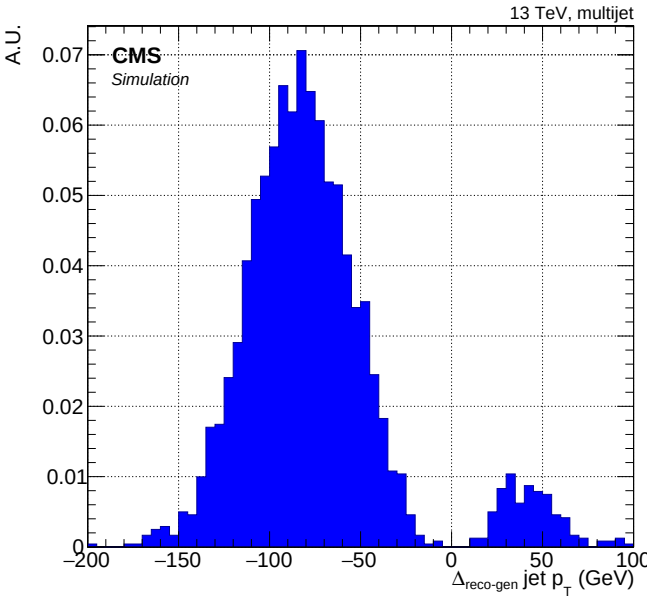


FIGURE 4.6: Difference between the reconstructed and generated p_T for the jet with the largest difference in absolute value. The plot has been produced using a multijet simulation with 500 GeV < genHT < 700 GeV and selecting the events with 80 GeV < MET < 120 GeV.

4.4.1 Multijet simulation

The multijet simulation has been produced in bins of the sum of the transverse parton momenta (genHT). The statistical power of the soft bins is too low: the bin with 500 GeV < genHT < 700 GeV, that gives the largest contribution in the signal region, has a cross-section of $\sim$ 32.1 nb and it has been generated with about 20 millions of events, corresponding to an equivalent luminosity of about 0.58 fb^{-1} that is too low compared with the 2.32 fb^{-1} of data collected. In the following, I describe a technique that I have developed to increase the statistical power of the multijet simulation in the large MET region, without generating new events.

In multijet events, fake MET is created by the mismeasurement of the transverse momentum of a jet or by the emission of a high-p_T neutrino originating from a hadron decay, as already discussed in Sect.4.2.5. Figure 4.6 shows the distribution of the largest difference in absolute value between the reconstructed and generated jet p_T in the event. The plot has been produced using simulated multijet events with 500 GeV < genHT < 700 GeV and selecting the events with reconstructed 80 GeV < MET < 120 GeV. The plot shows that large part of those events contains a jet whose energy is largely underestimated.

The technique used to increase the statical power of the multijet simulation consists in a smearing of the p_T resolution of a jet, in order to increase the MET. In detail, we used the following steps.

1. For each multijet simulation bin, the distribution of the ratio of the difference between the reconstructed and generated p_T for the jet with the largest difference in absolute value (see Fig. 4.6) is obtained, for events having $80\,\text{GeV} < \text{MET} < 120\,\text{GeV}$.

2. In each event, the reconstructed energy of the jet with the largest difference in absolute value between the reconstructed and generated p_T is smeared according to the distribution obtained in the step above (Fig. 4.6). Overestimated jets are allowed to become underestimated, and vice versa. This step is repeated until the MET is above 120 GeV.

3. All other variables of the events, like MHT and HT, are recomputed using the new jet p_T value.

4. The events are weighted so that the genHT distribution of the new events matches the distribution of original events with $\text{MET} > 120\,\text{GeV}$.

5. A correction is applied to emulate the trigger efficiency effect as a function of min(MET,MHT) (see Sect. 4.6).

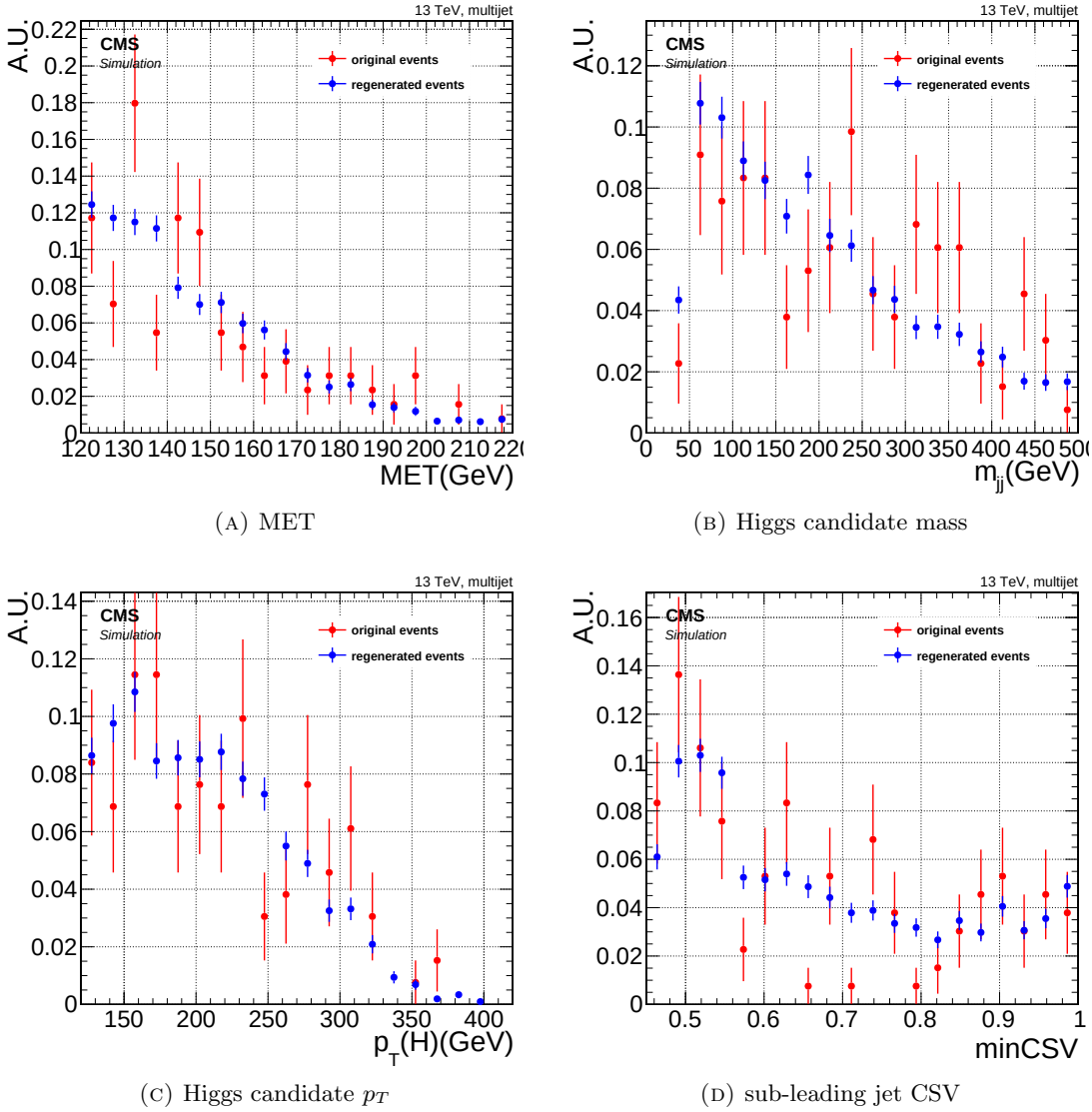
Figure 4.7 is a comparison of distributions obtained using the original multijet simulated events with $\text{MET} > 120\,\text{GeV}$ and the events produced with the technique described above. The new sample has a statical power significantly higher, and the two samples are compatible. The equivalent luminosity obtained with the regenerated sample in the multijet bin $500\,\text{GeV} < \text{genHT} < 700\,\text{GeV}$ is $10.1\,\text{fb}^{-1}$. The new sample has been also validated later (see Sect. 4.23), comparing the simulation with data.

4.5 Event reconstruction

4.5.1 Tracks and primary vertex

The tracks and primary vertices are reconstructed with the algorithms already described in Sect. 2.3.1 and 2.3.2. The simulations have been generated with a pile-up distribution roughly matching the data conditions. The small residual discrepancies have been eliminated reweighting events in simulations depending on the number of generated primary vertices. Figure 4.8 shows the distribution of the number of primary vertices reconstructed in the $t\bar{t}$-enriched control region, in data and simulation.

During the Run 1, the signal vertex was chosen as the primary vertex with the largest $\sum_{\text{tracks}} p_T^2$. I collaborated to improve this algorithm during the Long Shutdown 1. The idea is to choose the vertex depending on the $\sum p_T^2$ of the elementary particles produced in the collision (quarks, gluons, charged leptons, photons, and neutrinos). To do that,



(A) MET

(B) Higgs candidate mass

(C) Higgs candidate p_T

(D) sub-leading jet CSV

FIGURE 4.7: Comparison of the MET (A), Higgs candidate mass (B) and p_T (C), and sub-leading jet CSV (D) distribution between the original simulated multijet events (red points) and the ones regenerated with the technique described in the text (blue points). The selection MET > 120 GeV has been applied to both samples.

the $\sum p_T^2$ is evaluated for each primary vertex using, instead of tracks, the isolated charged leptons, the track-MET, and the track-jets. Figure 4.9 shows the performances of the two algorithms obtained in a signal simulation with pile-up forty and with pile-up distributed as the 2015 data.

4.5.2 Jets, MET, and leptons

Jets (Sect. 2.3.3) are reconstructed using the PF objects clustered with the anti-k_T algorithm with a distance parameter $R = 0.4$. Charged hadrons originating from the pile-up vertices are not included in the jet energy, in order to mitigate the pile-up effects.

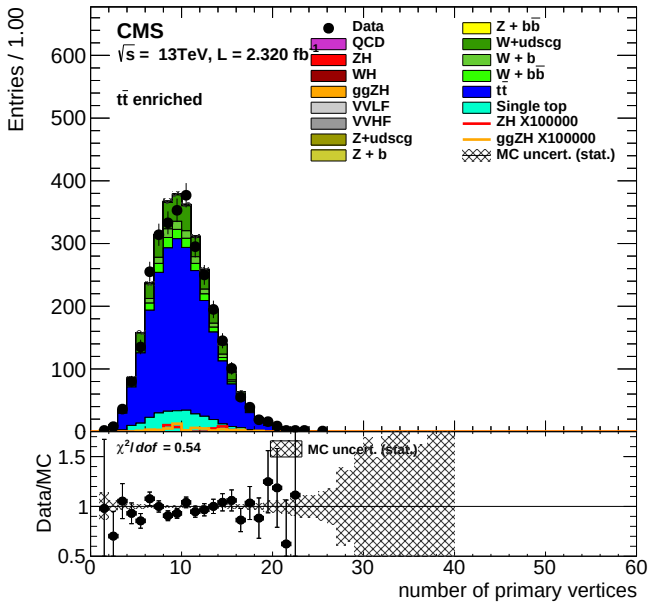


FIGURE 4.8: Number of primary vertices reconstructed in the $t\bar{t}$-enriched control region. The simulations have been normalized to match the data yield.

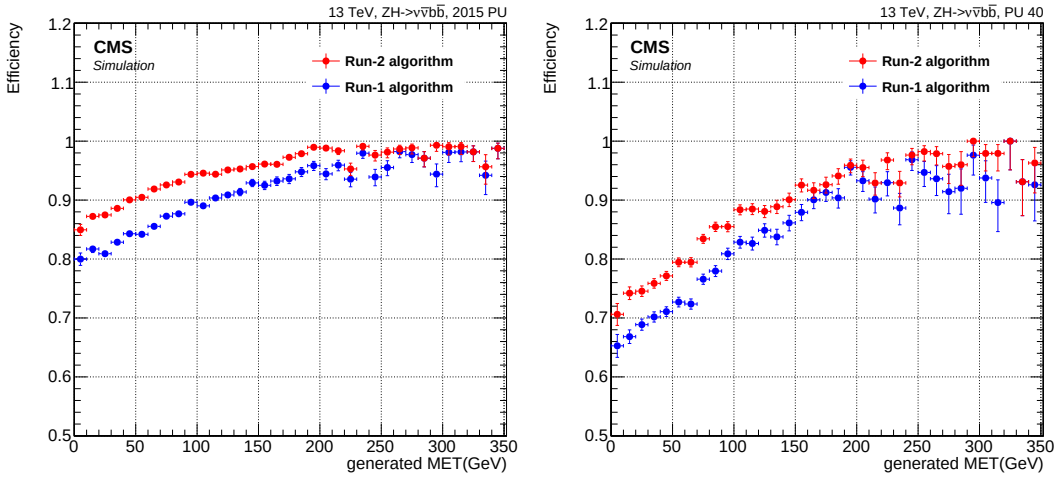


FIGURE 4.9: Efficiency to reconstruct the primary vertex within 500 μm from a collision producing the ZH signal. The plots have been obtained using simulation with pile-up distributed as the 2015 data (left) and with pile-up forty (right).

The jet energy is calibrated using corrections that make the jet energy response flat as a function of p_T, η, and number of primary vertices. The jet energy resolution obtained in simulation has been smeared in order to match the resolution measured in data.

In order to remove fake jets, jets are required to have the neutral hadron, neutral electromagnetic, and charged electromagnetic fractions less than 99% and containing at least two PF constituents and one charged hadron. The requirements on the charged components are applied only for jets within the tracker acceptance $|\eta| < 2.4$. To remove jets produced in pile-up collisions, a multivariate discriminant that combines information like the number of tracks, the jet energy components and the shape of the jets, is used. In this analysis, we use a pile-up identification working point with an efficiency of $\sim 90\%$ and a fake rate of $\sim 10\%$ for jets with $|\eta| < 2.5$ and $30\,\text{GeV} < p_T < 50\,\text{GeV}$.

The **MET** (Sect. 2.3.4) is reconstructed as the negative transverse vectorial sum of the PF objects. The MET is evaluated using the jet energy corrections improving the MET resolution. The **MHT** is defined as the negative transverse vectorial sum of the PF jets having $p_T > 30\,\text{GeV}$ and $|\eta| < 2.4$.

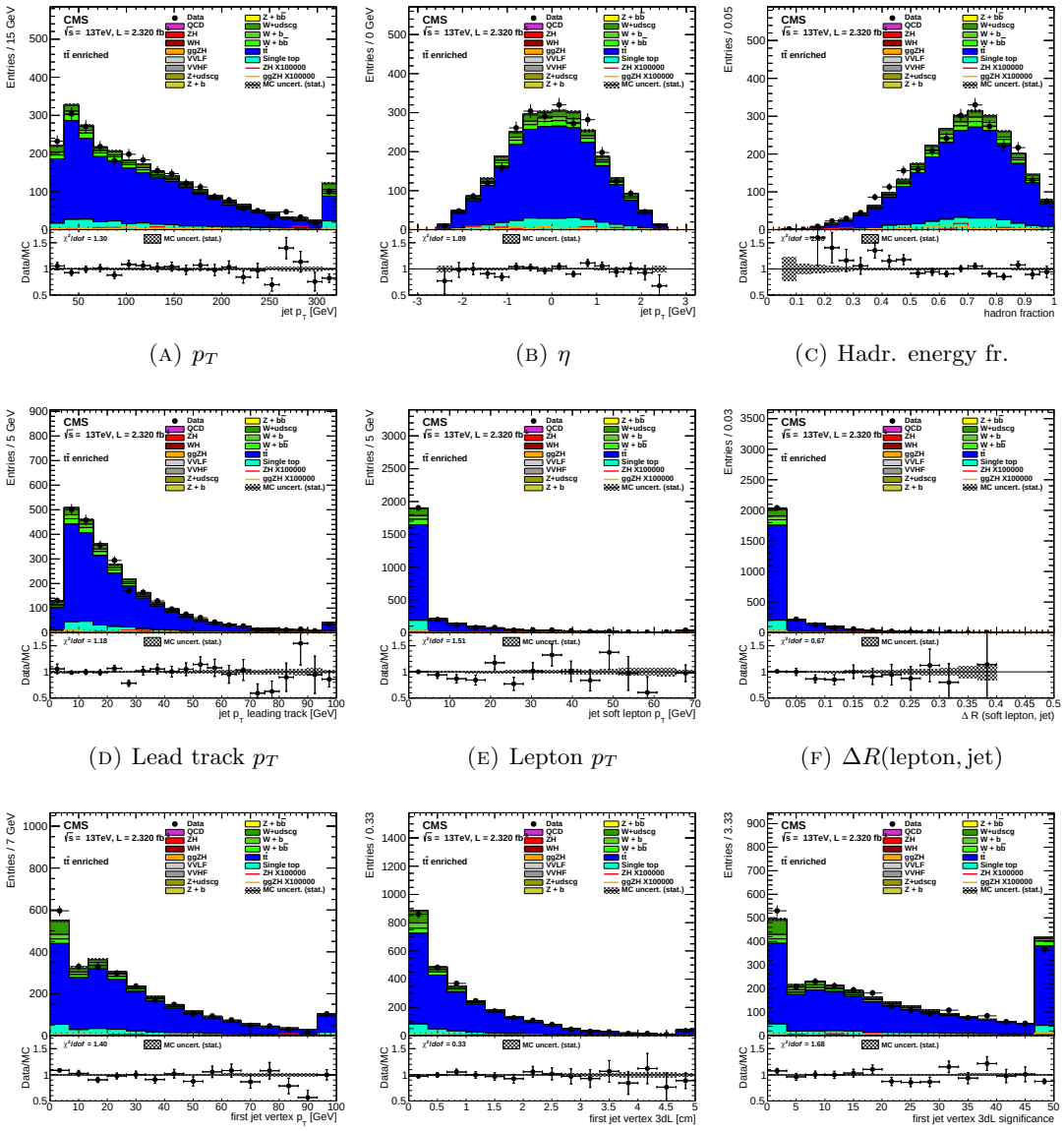
The **isolated charged leptons** (Sect 2.3.5 and 2.3.6) are required to have $p_T > 10\,\text{GeV}$ and an isolation of $\text{iso}_{\text{rel}} < 0.4$, defined as the sum of the PF objects, excluding the lepton and tracks originating from pile-up vertices, within a cone of $\Delta R < 0.4$, for muons, and $\Delta R < 0.3$, for electrons.

b-jet energy regression

About the 40% of B hadrons produced in the hadronization of b quarks decays with a neutrino in the final state. As the energy of the neutrino is not detected, the average jet energy response of b jets is lower with respect to uds jets, and the b-jet energy resolution is worse, as shown on the left side of Fig. 4.11. In attempt to recover such energy, a jet energy regression has been trained on a sample of simulated b jets, exploiting a few variables:

- p_T, η, transverse mass, hadronic energy fraction, electromagnetic energy fraction of the jet;
- p_T of the leading track within the jet;
- p_T, p_T relative to the jet axis, and ΔR of the lepton within the jet (if any);
- number of tracks, p_T, mass, 3D-impact parameter and 3D-impact parameter significance of the secondary vertex (if any);

Figure 4.10 shows the distribution of the variables used in the regression for data and simulation in a $t\bar{t}$-enriched control region.

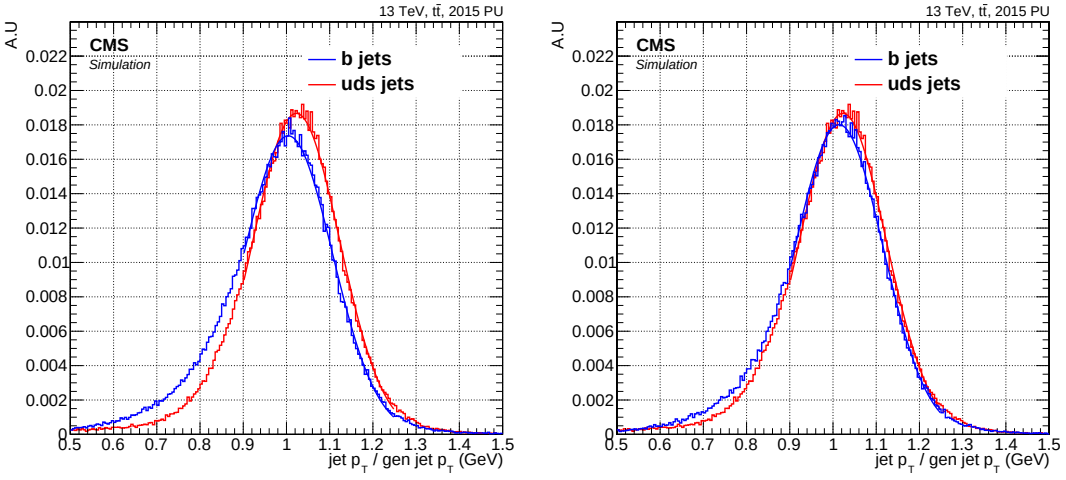


(A) p_T (B) η (C) Hadr. energy fr.

(D) Lead track p_T (E) Lepton p_T (F) $\Delta R(\text{lepton}, \text{jet})$

(G) Secondary vertex p_T (H) Sec. vertex 3D impact parameter (I) Sec. vertex 3D impact par. sign.

FIGURE 4.10: Distributions of the main variables used in the b-jet energy regression for the most b-tagged jets in the event. The events have been selected in the $t\bar{t}$-enriched control region.



(A) Baseline (B) Regressed

FIGURE 4.11: Jet energy response for *uds* (red) and b jets (blue) before (left) and after (right) the application of the b-jet energy regression. The regression has been applied only to the b jets. The energy response has been computed as the ratio between the reconstructed and the generated jets (including neutrinos), for jets with $p_T > 100\,\text{GeV}$ and $|\eta| < 2.4$. The mean values of the distributions are $\mu_b = 0.980$, $\mu_{\text{light}} = 1.011$, and $\mu_b^{\text{reg}} = 0.996$. The root mean squares are $\text{RMS}_b = 0.133$, $\text{RMS}_{\text{light}} = 0.122$, and $\text{RMS}_b^{\text{reg}} = 0.128$.

The right part of Fig. 4.11 shows the jet energy response after the regression. As expected, the mean b-jet energy response is closer to 1 and the root mean square is smaller with respect to the b-jet distribution before the regression.

The effects of the b-jet energy regression on the signal are shown in Fig. 4.12: the Higgs mass peak is reconstructed with a mean closer to 125 GeV and with a smaller standard deviation. The b-jet energy regression reduces the relative uncertainty (σ/μ) of $\sim 15\%$.

4.5.3 Identification of b jets

The b **tagging** (Sect. 2.3.8) plays a key role in the search for the $H \to bb$. The b-tagging algorithm used in this analysis is the combined secondary vertex v2 with the inclusive vertex finder (abbreviated to CSV in the rest of this chapter).

The CSV distributions are not perfectly reproduced by simulations and, for this reason, the simulations have been reweighted with the b-tagging scale factors, described in Sect. 2.3.8. Such scale factors depend on p_T, $|\eta|$, flavor and CSV, of the jets contained in each simulated event and they are obtained in $Z \to \ell\ell$+jets and fully leptonic $t\bar{t}$ control regions.

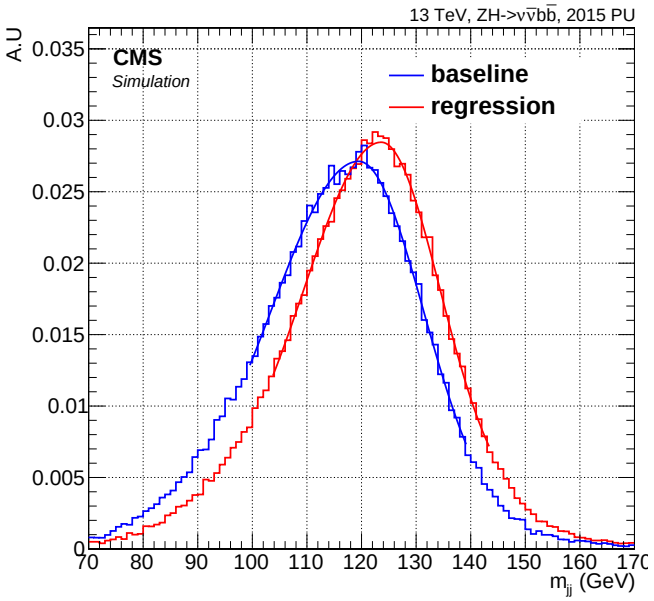


FIGURE 4.12: Invariant mass of the Higgs jets in $ZH \to \nu\nu bb$ simulation, before (blue) and after (right) the b-jet energy regression. The baseline m_{jj} distribution has been fitted with an asymmetric Gaussian having $\mu = 119.5\,\text{GeV}$, $\sigma_{\text{left}} = 16.0\,\text{GeV}$, and $\sigma_{\text{right}} = 11.9\,\text{GeV}$, and the regressed m_{jj} distribution with an asymmetric Gaussian having $\mu = 123.6\,\text{GeV}$, $\sigma_{\text{left}} = 14.2\,\text{GeV}$, and $\sigma_{\text{right}} = 10.8\,\text{GeV}$.

4.5.4 Higgs boson reconstruction

The Higgs boson candidate can be reconstructed using different criteria. Figure 4.13 shows the fraction of Higgs boson candidates selected correctly in a simulated signal sample, as a function of MET, for different selection criteria of the the jets from the decay $H \to bb$: the two most b-tagged jets, the two leading jets and the jet pair with the largest p_T. The most efficient method is to choose the two most b-tagged jets as Higgs candidate, and therefore it has been chosen in this analysis.

4.6 Triggers

During 2015, the maximum instantaneous luminosity of LHC was about $5 \cdot 10^{33}\,\text{cm}^{-2}\text{s}^{-1}$. At this luminosity, the unprescaled MET trigger with the lowest threshold was `HLT PFMET90 PFMHT90 IDTight` as described in Sect. 3.6.

The `HLT PFMET90 PFMHT90 IDTight` trigger cuts both on MET and MHT at the same threshold. For this reason, the trigger efficiency has been studied as a function of the offline min(MET,MHT). The pure MET trigger `HLT PFMET170` has been used as well, in order to recover a possible small inefficiency due to the online MHT cut in the large MET region.

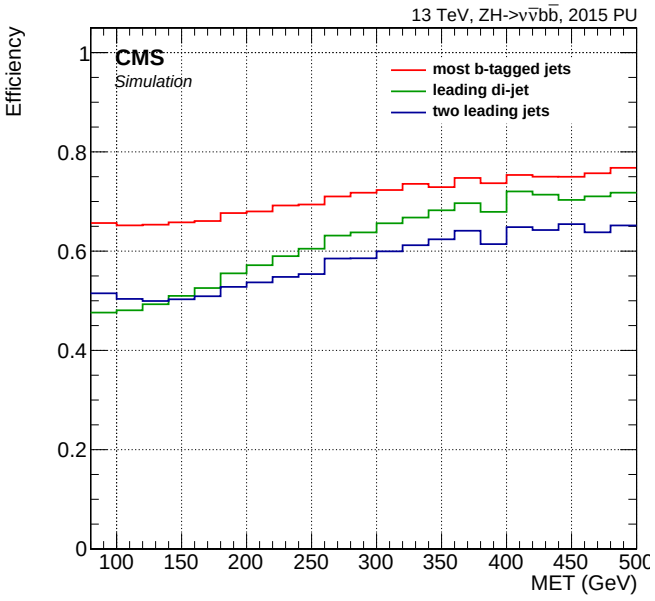
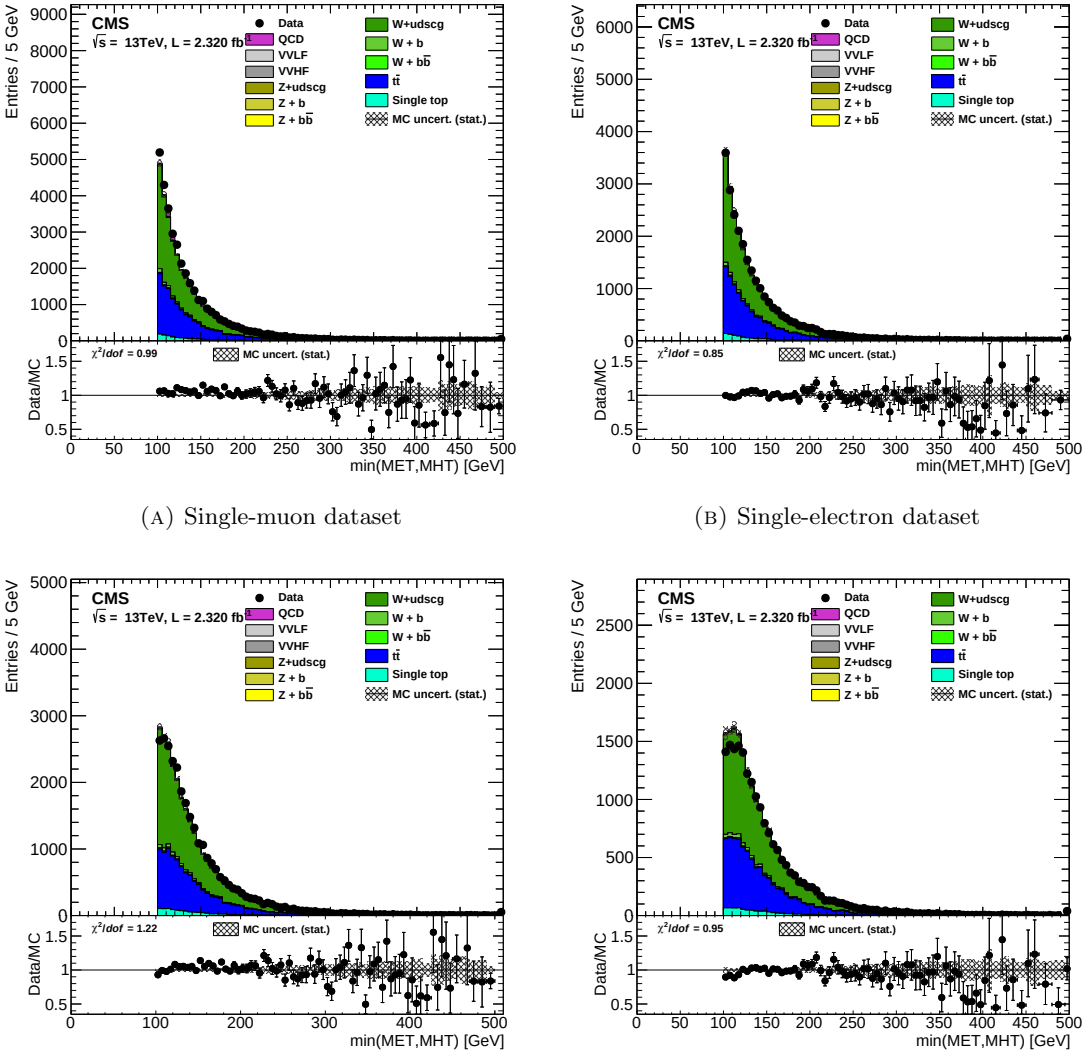


FIGURE 4.13: Efficiency of selecting correctly the jets from the decay of the Higgs boson in $b\bar{b}$ as a function of the MET, for a $ZH \to \nu\nu bb$ simulation. The plot shows the performance of selecting the Higgs jets as the two most b-tagged jets (red line), the jet pair with the largest p_T (green line), and the two leading jets (blue line). The plots have been obtained requiring two b jets with CSV > 0.46 and that the two b quarks originating from the Higgs decay have $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$.

The overall trigger efficiency has been measured using the data collected by the single-muon and single-electron triggers. The top plots in Figure 4.14 compare the min(MET,MHT) distributions obtained in data and simulation in the single-muon (left) and single-electron (right) datasets. The bottom plots have been obtained applying the same selections of the top plots with the addition of the requirement of the MET triggers.

The ratios between the bottom and top plots give the trigger efficiency curves. Figure 4.15 shows the efficiencies curves obtained independently in data and simulation (top plots) and the ratio between them (bottom plots). It shows that at min(MET,MHT) $\sim 120\,\text{GeV}$ the trigger efficiency is about 80% and for min(MET,MHT) $> 150\,\text{GeV}$ it is greater than 95%. The efficiency evaluated in data and in simulation gives similar results: for min(MET,MHT) $> 120\,\text{GeV}$ the discrepancy between the real and simulated trigger efficiency is lower than 5%.

The ratios between the trigger efficiency measured in data and in simulations have been fitted with error functions. The function fitted in the single-muon dataset (bottom-left plot in Fig. 4.15) are applied to the simulation as a trigger efficiency correction. The upper and lower curves shown in the plot have been taken as a systematic uncertainty. Figure 4.16 shows the distribution of min(MET,MHT), MET and MHT, in the single-electron dataset after applying both the MET trigger request and the trigger efficiency correction obtained in the single-muon dataset. The plots show that the



(A) Single-muon dataset

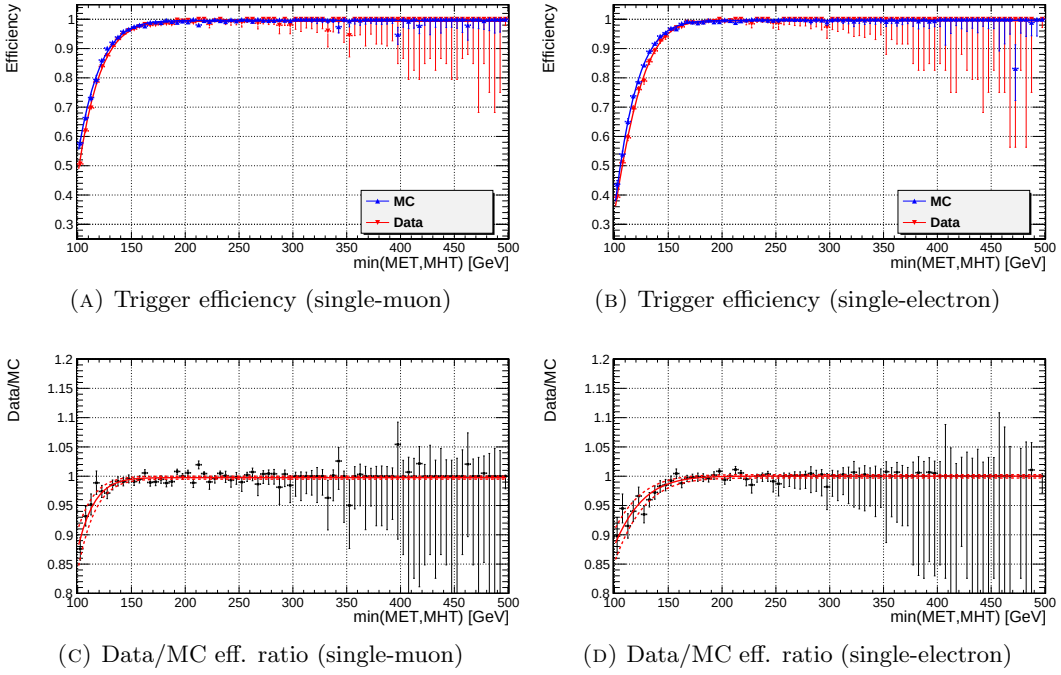
(B) Single-electron dataset

(C) Single-muon dataset & MET triggers

(D) Single-electron dataset & MET triggers

FIGURE 4.14: Distributions of min(MET,MHT) for the data and simulation in the single-electron (A) and single-muon (B) dataset. The bottom plots, (C) and (D), have been obtained applying the same selection of (A) and (B), respectively, with the addition of the trigger request `HLT PFMET90 PFMHT90 IDTight` or `HLT PFMET170`. A preselection of min(MET,MHT) > 100 GeV has been applied in all the plots.

trigger corrections work properly, as no significant data/simulation discrepancy appears in the min(MET,MHT) distribution. Moreover, the good data/simulation agreement in the MET and MHT distributions demonstrates that even if the triggers select events using two variables (MET and MHT), it is sufficient to apply the trigger correction as a function of only one variable (min(MET,MHT)).



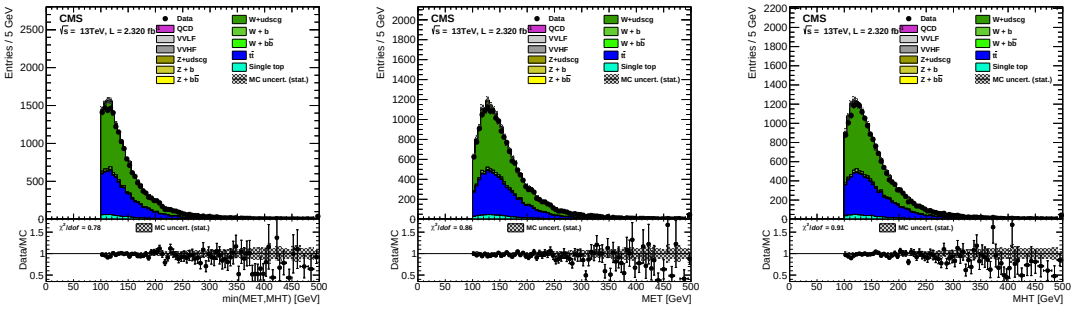
(A) Trigger efficiency (single-muon)

(B) Trigger efficiency (single-electron)

(C) Data/MC eff. ratio (single-muon)

(D) Data/MC eff. ratio (single-electron)

FIGURE 4.15: Trigger efficiency curves as a function of min(MET,MHT) in the single-muon (A) and single-electron dataset (B) for data (red) and simulation (blue). The efficiencies have been computed as the ratio of the plots shown in Fig. 4.14. Bottom: ratios between trigger efficiencies in data and simulation for the single-muon (C) and single-electron (D) dataset. The plots have been fitted with an error function. In the ratio plots, the 68% confidence band of the fit is shown as dashed line.

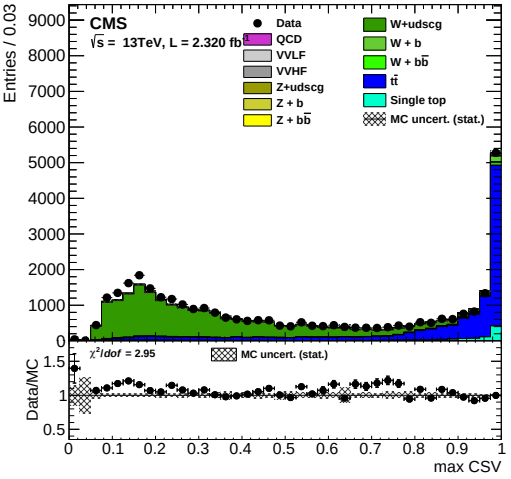


(A) min(MET,MHT)

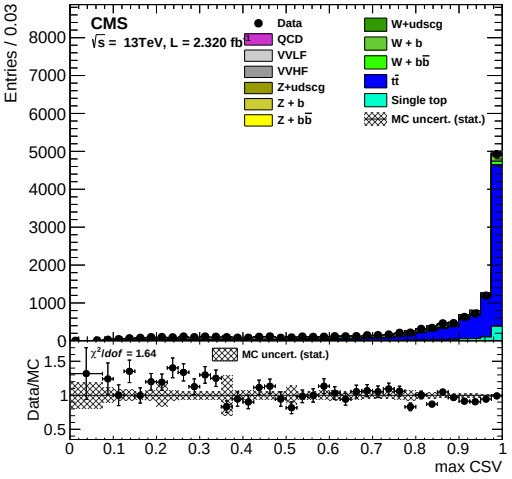
(B) MET

(C) MHT

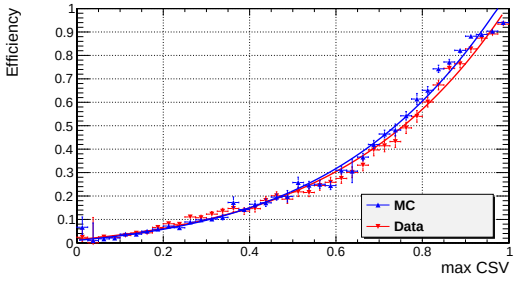
FIGURE 4.16: Distributions of min(MET,MHT), MET and MHT, for data and simulation. The plots have been obtained using the single-electron dataset, requiring the trigger `HLT PFMET90 PFMHT90 IDTight` or `HLT PFMET170`, and after applying the trigger corrections obtained from the single-muon dataset. The trigger correction function used is the fit of the bottom-left plot of Fig.4.15.



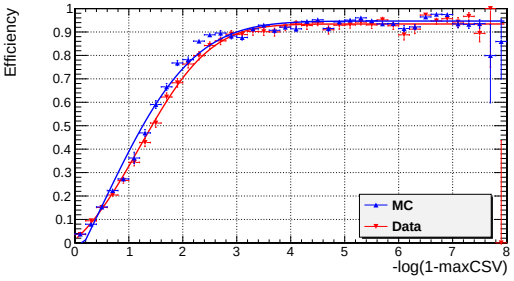
(A) $\mathrm{CSV}_{\mathrm{max}}$ (inclusive)



(B) $\mathrm{CSV}_{\mathrm{max}}$ (after online b tagging)



(C) Online b tag efficiency vs $\mathrm{CSV}_{\mathrm{max}}$



(D) Onl. b tag eff. vs $-\log(1-\mathrm{CSV}_{\mathrm{max}})$

FIGURE 4.17: Top, maximum jet CSV distribution as measured offline for the single-muon dataset after having applied the MET triggers without (A) and with (B) the request of the online b tagging. On the bottom left, the online b tagging efficiency curves for data (red) and simulation (blue) as a function of the maximum jet CSV (C). In order to visualize better the online b tagging efficiency in the region at $\mathrm{CSV}_{\mathrm{max}} \sim 1$, the plot (C) has been also plotted as a function of $-\log(1-\mathrm{CSV}_{\mathrm{max}})$ (D).

4.6.1 Trigger outlook for 2016

In the next few years, LHC will increase the instantaneous luminosity up to $\sim 2 \cdot 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$ and the rate of the `HLT PFMET90 PFMHT90 IDTight` will be too high to stay unprescaled. As discussed in Sect.3.6, we have prepared a trigger that exploits the online b tagging, in order to keep a low MET threshold even during the high luminosity data taking. The trigger `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72` was already running in 2015 and the trigger performance has been measured using single muon and single electron datasets. Figure 4.17 shows the online b-tagging performance of this trigger during 2015. The online b tagging looks well reproduced by simulation.

The online b tagging efficiency is about 97.8% for signal events having two jets passing, respectively, the loose and medium b tagging working point ($\mathrm{CSV}_{\mathrm{loose}} > 0.460$,

$\text{CSV}_{\text{medium}} > 0.800$). This b tagging selection is the same used in the signal region, defined later in Sect. 4.7.

As already discussed in Sect. 3.6, the online b tagging reduces the rate of a factor ~ 5 and hence the rate of `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72` is sustainable even in the high luminosity scenario ($\mathcal{L} = 1.4 \cdot 10^{34}\,\text{cm}^{-2}\text{s}^{-1}$), when it is expected to be $\sim 28\,\text{Hz}$.

4.7 Event selection

4.7.1 Preselection

A preselection is applied to the signal and all control regions. The events are required to be triggered by `HLT PFMET90 PFMHT90 IDTight` or `HLT PFMET170` and to have: MET $> 120\,\text{GeV}$, MHT $> 120\,\text{GeV}$, at least two jets with CSV > 0.460 (loose working point), $p_T > 20\,\text{GeV}$ and $|\eta| < 2.4$, $p_T(\text{Higgs candidate}) > 120\,\text{GeV}$. $m(\text{Higgs candidate}) < 500\,\text{GeV}$, and $\Delta\varphi(\text{Higgs candidate}, \text{MET}) > \pi/2$.

Beam-halo noise cleaning

The leading jet in the event is explicitly required to have $|\eta| < 2.4$ and a charged hadron energy fraction greater than 10%. This cut removes the machine-induced background due to the beam halo, i.e. the set of particles generated by the beam interaction with the residual gas in the vacuum chamber or with a LHC collimator or with the beam pipe. When a beam halo particle interacts with the CMS detectors, it may release a large amount of energy in the calorimeters, creating a very energetic fake jet and fake MET. The main characteristic of such fake jet is a very small charged hadron fraction, because there are no or very few tracks pointing towards the direction of the fake jet. The beam-halo particles are mainly spread along the radial direction of the LHC, corresponding to $\varphi \sim 0$ or $\varphi \sim \pi$ in the CMS coordinates.

Figure 4.18 shows the charged hadron fraction distribution of the leading jet, in a beam-halo noise cleaned and enriched region. The beam-halo cleaned region is defined as $0.3 < \varphi(\text{lead. jet}) < \pi - 0.3$ and the beam-halo enriched region as the complementary. As expected, the charged hadron fraction distribution shows a clear peak in 0 only in the beam-halo enriched region.

4.7.2 Anti-QCD cuts

In order to reject the multijet background, two sets of cuts are used. The **anti-QCD loose** cut has been defined requiring:

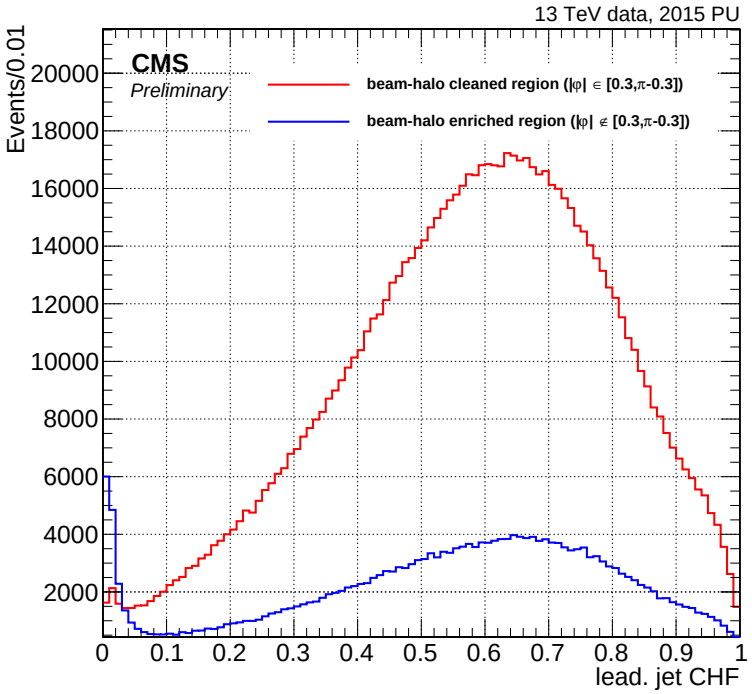


FIGURE 4.18: Charged hadron fraction of the leading jet in beam-halo cleaned (red) and enriched (blue) control regions. The beam-halo cleaned region requires $0.3 < \varphi(\text{lead. jet}) < \pi - 0.3$ and the beam-halo enriched region is the complementary. The plots have been produced using the data collected in 2015 selecting events passing the MET triggers and offline MET and MHT $> 120\,\text{GeV}$. As expected, a large fraction of events in the beam-halo enriched region has the leading jet with a charged hadron fraction lower than 0.1.

- the leading jet not pointing towards the hot ECAL cells as mapped in Fig. 4.4c and 4.4d, if it has $\Delta\varphi(\text{jet}, \text{MET}) > \pi - 0.4$;
- the jets with $p_T > 30\,\text{GeV}$ not pointing towards the dead ECAL cells as mapped in Fig. 4.4a and 4.4b, if they have $\Delta\varphi(\text{jet}, \text{MET}) < 0.4$;
- both Higgs candidate jets with $\Delta\varphi(\text{jet}, \text{MET}) > 0.4$;
- two leading jets with $\Delta\varphi(\text{jet}, \text{MET}) > 0.4$;

The anti-QCD loose cut rejects $\sim 81\%$ of multijet events having $500\,\text{GeV} < \text{genHT} < 700\,\text{GeV}$, with a signal efficiency in the signal region of $\sim 97\%$.

In the **anti-QCD tight** cut, we extend the $\Delta\varphi(\text{jet}, \text{MET}) > 0.4$ cut to all jets with $p_T > 30\,\text{GeV}$. The cut rejects the $\sim 91\%$ of multijet events having $500\,\text{GeV} < \text{genHT} < 700\,\text{GeV}$, with a signal efficiency in the signal region of $\sim 93\%$.

In the signal region, only the anti-QCD loose cut is applied, in order to keep a high signal efficiency. The tight cut is used to increase the purity of the Z + light-jets and Z + b-jets, in the respective control regions.

4.7.3 Background and signal region

Six control regions and one signal region have been defined for the two bins of $p_T(Z)$, defined as min(MET,MHT): $120\,\text{GeV} < p_T(Z) < 150\,\text{GeV}$ and $p_T(Z) > 150\,\text{GeV}$. The cuts used to define the signal and control regions are reported in Tab. 4.1.

- The signal region is defined requiring the $Z \to \nu\nu$ category, to be orthogonal with the other $VH(bb)$ searches. A lepton veto is applied and at least one medium b-tagged jet is required. A m_{jj} cut is used to separate the signal from the $Z+$b-jets control region. The multijet background is reduced applying the anti-QCD loose cut.

- The $Z+$b-jets control region is similar to the signal region, but inverting the m_{jj} cut. The anti-QCD tight cut is used to increase the $Z+$b-jets purity.

- The $Z+$light-jets control region is defined inverting the b-tagging cut and removing the m_{jj} cut. The remaining cuts are identical to the $Z+$b-jets control region.

- In the multijet control region, the anti-QCD tight cut is inverted. The loose anti-QCD cut is still applied in order to have the control region as close as possible to the signal region.

- The $t\bar{t}$ control region is defined by a medium b-tagged jet, at least one isolated lepton and four jets with $p_T > 30\,\text{GeV}$.

- The $W+$b-jets control region is made independent from the $t\bar{t}$ control region by inverting the four jets cut. In order to reduce the $t\bar{t}$ contamination, exactly one additional lepton is required and it is required to have a $\Delta\varphi(\text{lept.,MET}) < \pi/2$. This cut is useful in rejecting the fully leptonic decay of $t\bar{t}$. The cut $\Delta R_{jj} \leq \pi/3$ on the Higgs candidate jets is used to reduce the $t\bar{t}$ contamination as well. A m_{jj} veto is used to remove the possible $WH \to \ell\nu bb$ contribution into this control region.

- In the $W+$light-jets control region, the medium b-tagging cut is inverted and no m_{jj} veto is applied.

TABLE 4.1: Control and signal regions for the high and low $p_T(Z)$ bins. The thresholds reported in the $\min(p_T(H),p_T(Z))$ row in the bottom of the table are for the high $p_T(Z)$ bin only. The cuts for the low $p_T(Z)$ bin are reported in the subsequent row. The remaining rows are intended for both the high and low $p_T(Z)$ bins. The cuts are applied after the preselection described in Sect. 4.7.1.

Variable	$t\bar{t}$	W + light	W + b	multijet	Z + light	Z + b	Signal
V decay category	$W \to \ell\nu$	$W \to \ell\nu$	$W \to \ell\nu$	$Z \to \nu\nu$	$Z \to \nu\nu$	$Z \to \nu\nu$	$Z \to \nu\nu$
$N_{\text{lept.}}$	≥ 1	$= 1$	$= 1$	$= 0$	$= 0$	$= 0$	$= 0$
N_{jets}	≥ 4	≤ 3	≤ 3	–	–	–	–
$\Delta\varphi$(lept.,MET)	–	–	$\leq \pi/2$	–	–	–	–
ΔR_{jj}	–	–	$\leq \pi/4$	–	–	–	–
m_{jj}	–	–	$\notin [90-150]$	–	–	$\notin [90-150]$	$90-150$
CSV_{max}	> 0.800	< 0.800	> 0.800	–	< 0.800	> 0.800	> 0.800
$\text{anti-QCD}_{\text{loose}}$	–	–	–	yes	yes	yes	yes
$\text{anti-QCD}_{\text{tight}}$	–	–	–	inverted	yes	yes	–
$\min(p_T(H),p_T(Z))$	> 150	> 150	> 150	> 150	> 150	> 150	> 150
(low $p_T(Z)$ bin)	$120-150$	$120-150$	$120-150$	$120-150$	$120-150$	$120-150$	$120-150$

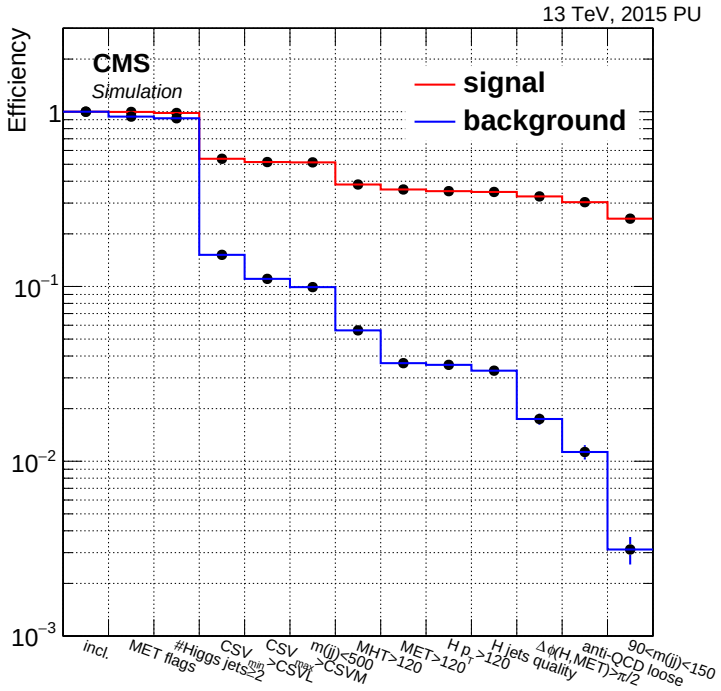


FIGURE 4.19: Efficiency and background reduction after each cut used in the signal region definition, for signal and overall background. The plot have been obtained requiring the MET triggers and $Z \to \nu\nu$ category as preselection.

Figure 4.19 shows the efficiency and the background reduction after each cut used in the signal region definition.

Data-driven background normalization

The normalizations of the $t\bar{t}$, W+0b−jets, W+1b−jets, W+2b−jets, Z+0b−jets, Z+1b−jets, Z+2b−jets, and multijets backgrounds are freely fitted in each $p_T(Z)$ bin during the signal extraction. In section 4.7.4 we describe the multivariate discriminant used to extract the signal. Preliminary background normalizations are needed for its training and to test the modeling of the variables used in the multivariate analysis.

The preliminary background normalizations have been fitted in the six control regions using the distributions of the minimum CSV among the Higgs candidate jets ($\mathrm{CSV_{min}}$). This variable has been chosen in order to increase the separation between V+1/2b−jets and V+0b−jets backgrounds within the control regions. The fit has been performed merging the two $p_T(Z)$ bins of the analysis and using a common scale factor for Z/W+1b−jets and Z/W+2b−jets, in order to avoid large statistical fluctuations. The fitted scale factors on the background normalizations are reported in Tab.4.2. Figure 4.20 and 4.21 show the fitted distributions, before and after the application of the preliminary scale factors. The preliminary scale factors have been found to be close to

unity for all processes with the exception of the $Z + b$-jets. The excess of the $Z + b$-jets scale factor is driven by the $Z + 1b$-jets background, as it will be shown and discussed in Sect. 4.10.

4.7.4 Multivariate discriminant

A Boosted Decision Tree (BDT) [82, 83] has been trained to separate signal from background in the signal regions of the two $p_T(Z)$ bins. The BDT exploits the following variables:

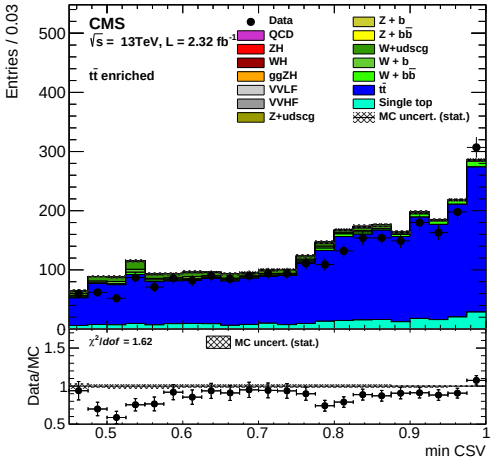
- p_T, η, $\Delta\eta(jj)$, $\Delta\varphi(jj)$ of the Higgs candidate;
- $\mathrm{CSV}_{\mathrm{min}}$,$\mathrm{CSV}_{\mathrm{max}}$, $p_{T,\mathrm{min}}$, $p_{T,\mathrm{max}}$ of the Higgs jets;
- $\mathrm{CSV}_{\mathrm{max}}$ and $p_{T,\mathrm{max}}$ of the additional jets (if any);
- $\Delta\varphi(\mathrm{Higgs}, \mathrm{MET})$;
- number of track jets with $p_T > 2\,\mathrm{GeV}$ $(R = 0.4)$ [2];
- minimum $\Delta\varphi(\mathrm{jet}, \mathrm{MET})$ among jets with $p_T > 30\,\mathrm{GeV}$.

The distribution of the most discriminating variables along with the BDT output are shown in Fig. 4.22 for the $Z + b$-jets control region. The plots of the multijet control region are shown Fig. 4.23. The data/simulation agreement validates the method used to increase the statistical power of the multijet simulation, described in Sect. 4.4.1. The plots of the remaining control regions are shown in Appendix C (Fig. C.1, C.2, C.3, and C.4).

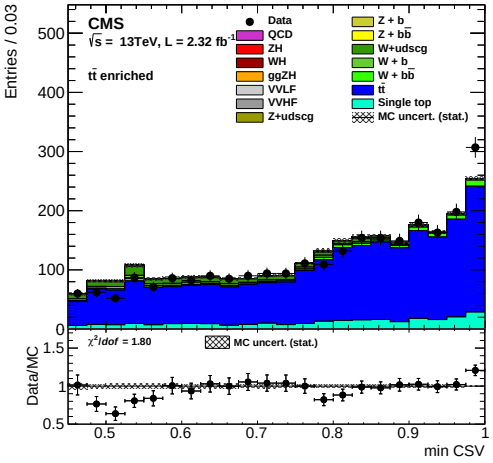
TABLE 4.2: Preliminary scale factors of the main background normalizations. The quoted error is statistical only.

Background	Scale factor
QCD multijet	1.32 ± 0.11
$t\bar{t}$	0.87 ± 0.02
W + light-jets	1.14 ± 0.06
W + b-jets	1.11 ± 0.24
Z + light-jets	1.24 ± 0.10
Z + b-jets	1.87 ± 0.14

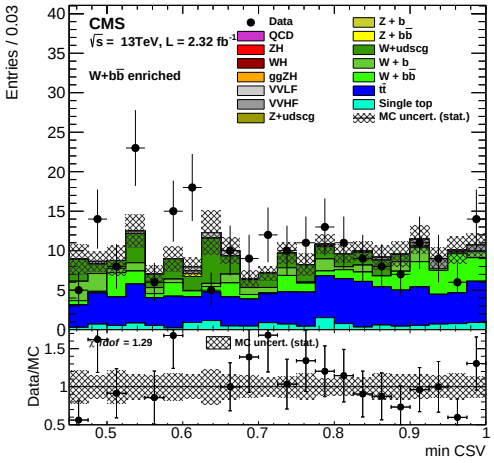
[2] This variable is useful to measure the soft-activity in the event. As in the $q\bar{q}ZH$ production there is no interaction with gluons, we expect to have a lower soft activity in the events. Moreover, since it is measured using track-jets, this variable is pile-up independent.



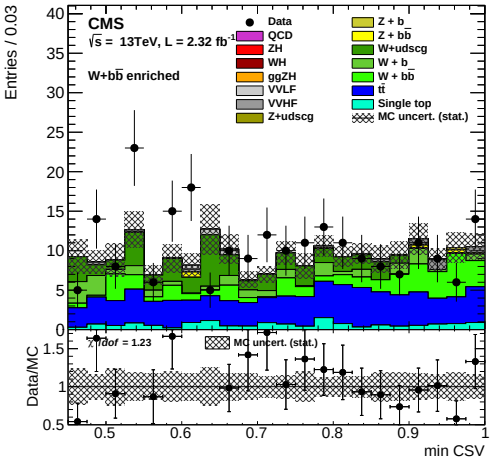
(A) $t\bar{t}$ c.r. - w/o scale factors



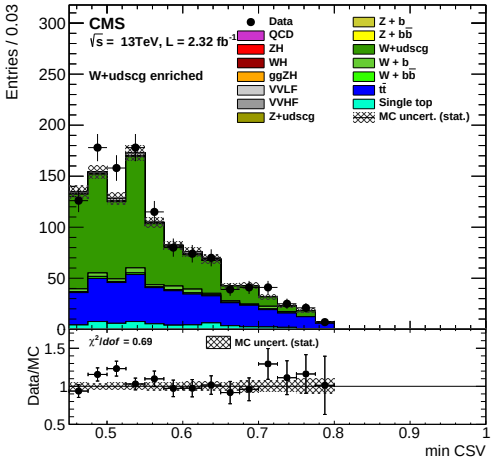
(B) $t\bar{t}$ c.r. - w/ scale factors



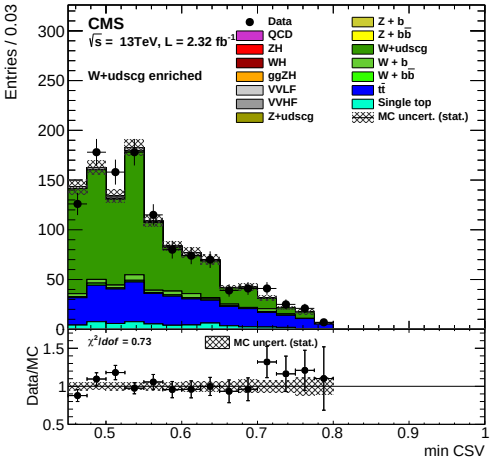
(C) W + b-jets c.r. - w/o scale factors



(D) W + b-jets c.r. - w/ scale factors

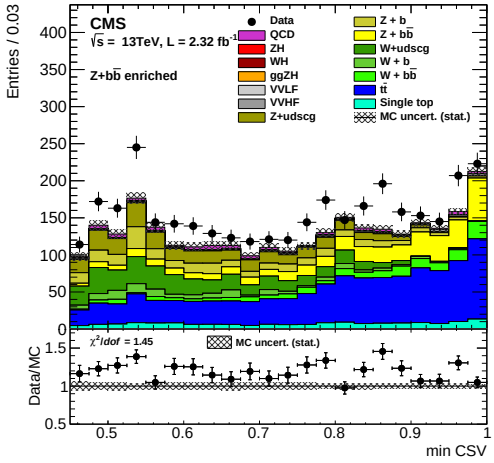


(E) W + light-jets c.r. - w/o scale factors

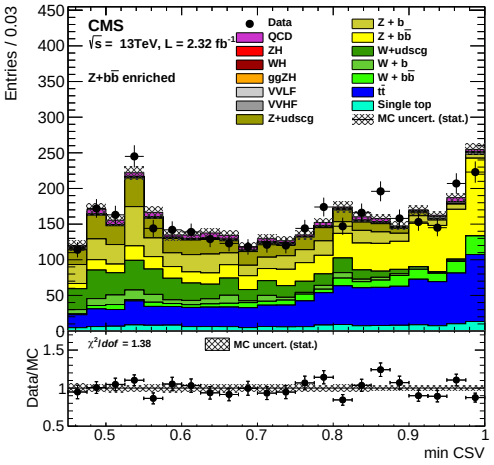


(F) W+light-jets c.r. - w/ scale factors

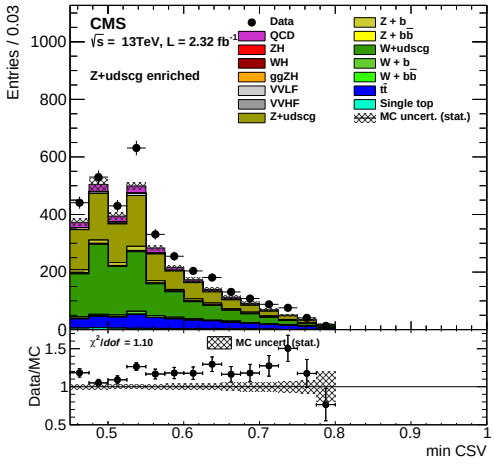
FIGURE 4.20: $\mathrm{CSV}_{\mathrm{min}}$ before (left) and after (right) the preliminary scale factor application. The plots have been obtained in the $t\bar{t}$ ((A),(B)), W + b-jets ((C),(D)), and W + light-jets ((E),(F)) control regions.



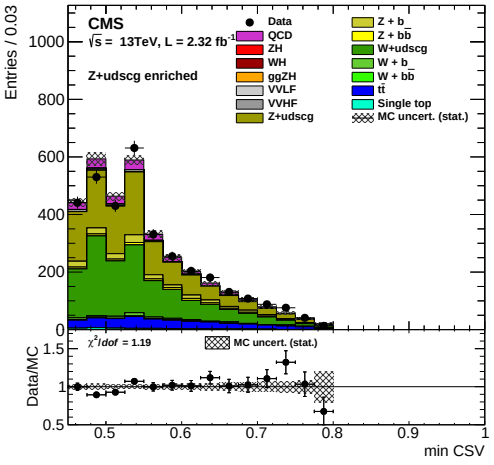
(A) Z + b-jets c.r. - w/o scale factors



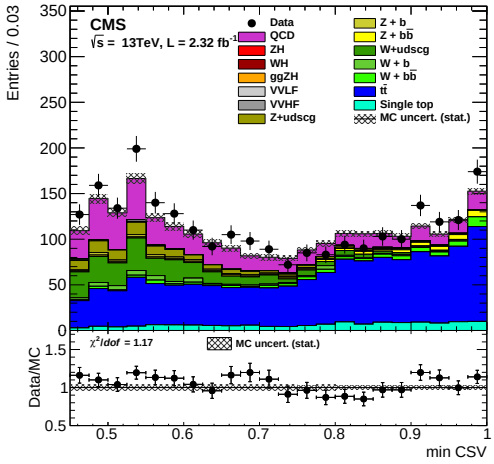
(B) Z + b-jets c.r. - w/ scale factors



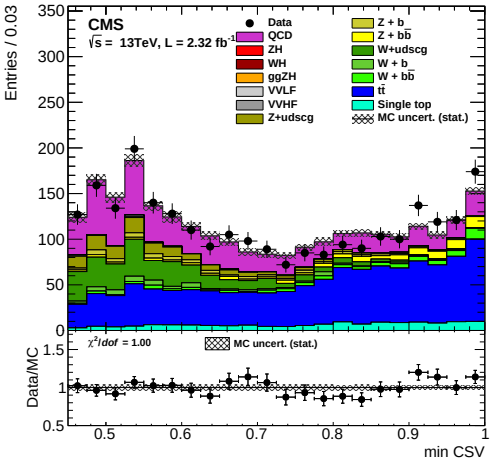
(C) Z + light-jets c.r. - w/o scale factors



(D) Z+light-jets c.r. - w/ scale factors

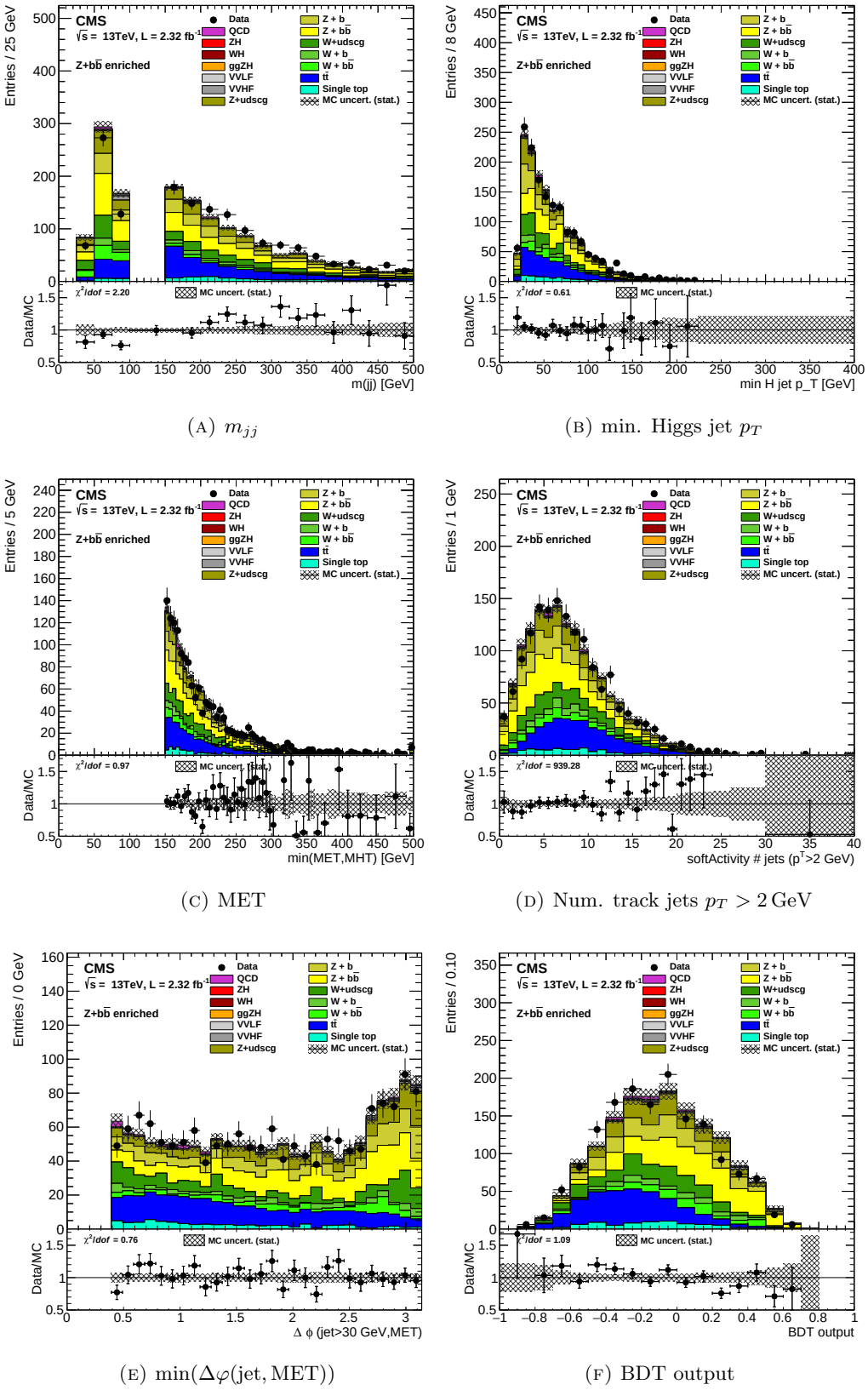


(E) Multijet c.r. - w/o scale factors



(F) Multijet c.r. - w/ scale factors

FIGURE 4.21: CSV_{min} before (left) and after (right) the preliminary scale factor application. The plots have been obtained in the Z + b-jets ((A),(B)), Z + light-jets ((C),(D)), and multijet ((E),(F)) control regions.

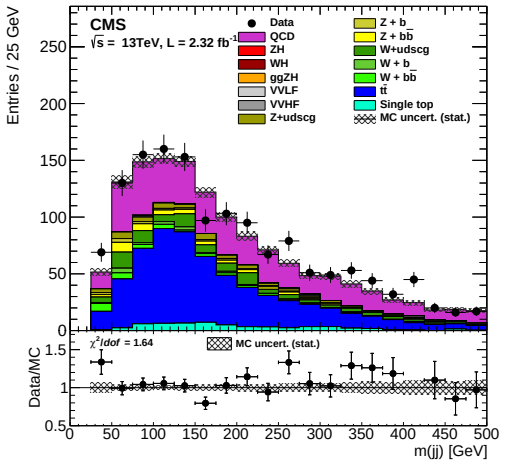


(A) m_{jj} (B) min. Higgs jet p_T

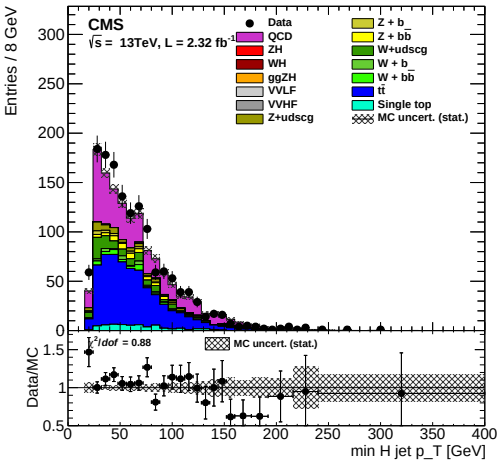
(C) MET (D) Num. track jets $p_T > 2\,\text{GeV}$

(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$ (F) BDT output

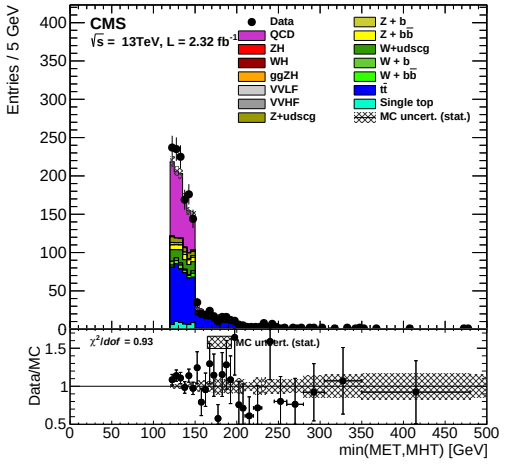
FIGURE 4.22: Distributions of some variables for the Z+b-jets control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



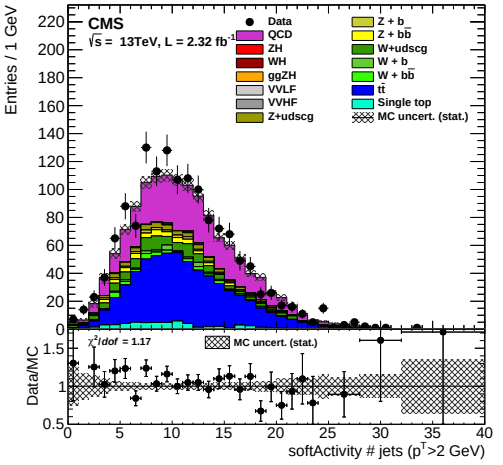
(A) m_{jj}



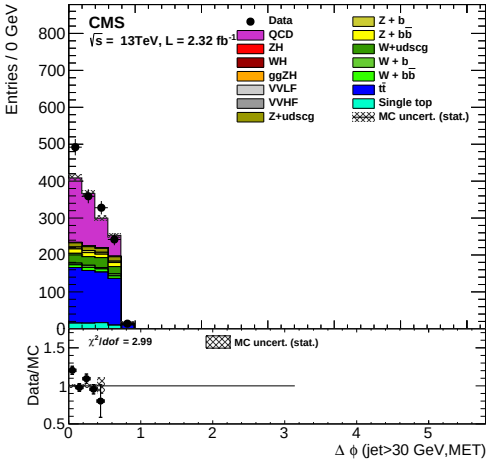
(B) min. Higgs jet p_T



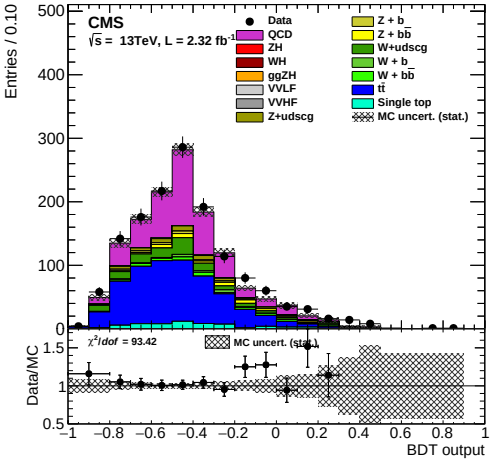
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$



(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$



(F) BDT output

FIGURE 4.23: Distributions of some variables for the multijet control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.

4.8 Systematics

The sensitivity of the search for the $ZH \to \nu\nu bb$ is largely dominated by the statistic uncertainty. The shape systematics that have been considered are following.

- The uncertainty on the b-tagging scale factors, used to reshape the CSV distributions (described in Sect.4.7.3). As this scale factors are obtained using data, their uncertainties are expected to decrease with the increasing of the integrated luminosity.
- The MC statistics error is taken as uncorrelated bin-by-bin shape uncertainty. It is significant for the soft production of the W/Z+jets.
- The uncertainty on the jet energy calibration and resolution.
- The uncertainty on the trigger efficiency corrections (see bottom-left plot in Fig. 4.15).
- The error related on the number of simultaneous proton-proton collision uncertainty has been estimated shifting this distribution of a factor $\pm 5\%$.
- The signal modeling (Poweheg vs aMC@NLO).

The dominating shape systematics are those related to the b-tagging reshaping and to the MC statistics. The main normalization systematics are the following.

- The luminosity uncertainty (4%) is propagated to the signal and to the not-data-driven background normalizations (single-top and di-boson).
- The effects of the parton distribution functions uncertainty are about 2% for the quark-quark initiated process ($q\bar{q}ZH$,WH,VV) and about 8% for the others ($ggZH$ and single top).
- The error on the cross-sections of the not data-driven backgrounds have been estimated in $\sim 15\%$.
- The uncertainty on the QCD scale changes the signal and the di-boson normalizations of $\sim 5\%$, for the $q\bar{q}$ initial state, and of $\sim 30\%$ for the gg initial state.
- The error due to the electroweak correction is about 5% for the signal.

4.9 Signal extraction

The test statistic and the treatment of the nuisance parameters follow the recommendations of the LHC Higgs Combination group [84, 85].

Each systematic uncertainty is described using a nuisance parameter θ_i having as probability distribution $p_i(\widetilde{\theta_i}|\theta_i)$ a Gaussian with $\mu = 0$ and $\sigma = 1$. The expected signal and background models are described depending on the parameters θ_i. As the multijet, Z + light-jets, $Z + b$-jets, W + light-jets, $W + b$-jets, and $t\bar{t}$ background normalizations are unconstrained, they are parametrized with nuisance parameters having flat $p_i(\widetilde{\theta_i}|\theta_i)$.

The global likelihood function is expressed as the product of likelihoods of the data in each BDT plot for signal regions and $\mathrm{CSV_{min}}$ plot for control regions times the product of the $p_i(\widetilde{\theta_i}|\theta_i)$:

$$L(\mathrm{data}, \widetilde{\theta}|\mu, \theta) = \prod_{n=1}^{N_p} L_n(\mathrm{data}_n|\mu, \theta) \times \prod_{i=1}^{N_\theta} p_i(\widetilde{\theta_i}|, \theta_i), \tag{4.1}$$

where N_p is the number of fitted plots (i.e. 2 $p_T(Z)$-bins $\times$ 7 regions = 14), N_θ is the number of nuisance parameters (about 300), and μ is a multiplying factor to the signal normalization ($\mu = 1$ corresponds to the SM).

In each plot, the likelihood is defined as:

$$L_n(\mathrm{data}|\mu, \theta) = \prod_{n=1}^{N_B} \mathrm{Poisson}(n_k|\mu \cdot s_k(\theta) + b_k(\theta)), \tag{4.2}$$

where s_k and b_k are, respectively, the expected signal and background yield in bin k, n_k is the observed yield, and N_B is the number of bins.

The exclusion limit on the Higgs boson hypothesis is set using a test statistic (q_μ), defined depending on μ:

$$q_\mu = \frac{L(\mathrm{data}, \widetilde{\theta}|\mu, \hat{\theta_\mu})}{L(\mathrm{data}, \widetilde{\theta}|\hat{\mu}, \hat{\theta})} \qquad \text{with } 0 \leq \hat{\mu} \leq \mu, \tag{4.3}$$

where $\hat{\theta_\mu}$ is the θ value maximizing the likelihood at the numerator given μ, and $\hat{\mu}$ and $\hat{\theta}$ are the values maximizing the likelihood at the denominator.

Finally, the CL_s [86–88] is defined as:

$$\mathrm{CL}_s(\mu) = \frac{\mathrm{CL}_{s+b}}{\mathrm{CL}_b} = \frac{P\left(q_\mu \geq q_\mu^{\mathrm{obs}}|\mu\right)}{P\left(q_\mu \geq q_\mu^{\mathrm{obs}}|0\right)} \tag{4.4}$$

4.10 Results

The final signal extraction is a simultaneous fit of the BDT distribution in the signal region and of the $\mathrm{CSV_{min}}$ distribution in the six control regions, for both the high and low $p_T(Z)$ bin. In the control regions, the $\mathrm{CSV_{min}}$ distribution has been chosen in order to increase the separation between the $V + 1b-$jets and $V + 2b-$jets backgrounds. The

fitted distributions are shown in Fig. 4.24 (control regions, high $p_T(Z)$ bin), Fig. 4.25 (control regions, low $p_T(Z)$ bin), and Fig. 4.26 (signal regions, high and low $p_T(Z)$ bin).

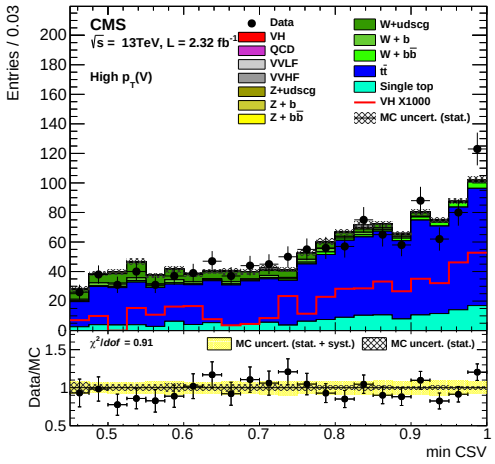
The observed 95% CL upper limit on the cross-section of $Z(\nu\nu)H(bb)$ is 3.97 times the SM prediction. The expected limit is 3.95. An excess of 0.52 standard deviations over the SM backgrounds is expected, but a slight under-fluctuation is found. The fitted signal strength is $\mu = -0.25^{+1.47}_{-1.37}$. This measurement is compatible with both the background only and SM Higgs boson hypotheses within one standard deviation. Figure 4.27 shows the likelihood scan for the parameter μ.

Combining the search for $Z(\nu\nu)H(bb)$ with the $W(\ell\nu)H(bb)$ and the $Z(\ell\ell)H(bb)$ channels, the analysis sensitivity increases significantly. Figure 4.28 shows the expected upper limit of the searches for $Z(\ell\ell)H(bb)$, $W(\ell\nu)H(bb)$, $Z(\nu\nu)H(bb)$, and their combination. The expected significance of the combination is 0.81 standard deviations and the expected uncertainty on the signal strength is 1.06.

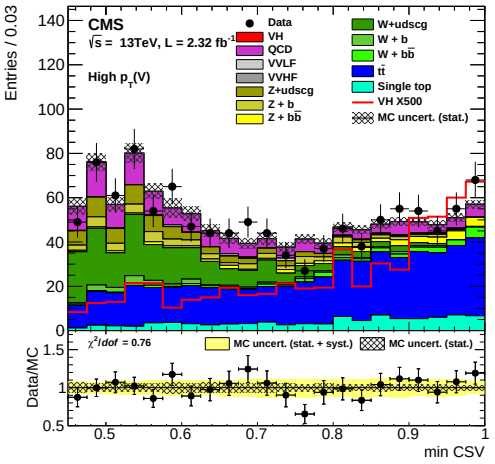
The scale factors of the background normalizations obtained in the final fit are shown in Table 4.3. All values have been found to be compatible with the preliminary scale factors showed in Table 4.2, except the $Z+2b-$jets that is compatible to one (the overall preliminary scale factor of $Z+b-$jets was 1.87). On the other hand, the $Z+1b-$jets scale factors have been found to be even larger the preliminary value. Large $Z+1b-$jets scale factors were already measured in the Run-1 $VH(bb)$ analysis [27], and they are due to a mismodeling in the generator parton shower of the process of gluon-splitting to b-quark pairs. The effect of the other nuisance parameters after the final fit is to increase the background normalizations of about 0% – 15%. Since this effect changes significantly among the regions and the backgrounds considered, it has not been included in Table 4.3.

TABLE 4.3: Scale factors of the main background normalizations obtained after the final fit. The quoted uncertainty includes both the statistic and systematic errors.

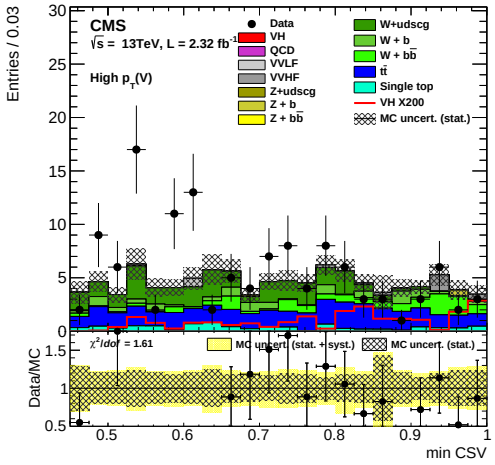
Background	High $p_T(Z)$	Low $p_T(Z)$
QCD multijet	0.95 ± 0.24	1.50 ± 0.12
$t\bar{t}$	0.86 ± 0.05	0.92 ± 0.05
$W+0$b-jets	1.13 ± 0.12	1.18 ± 0.12
$W+1$b-jets	1.83 ± 1.25	1.12 ± 1.06
$W+2$b-jets	1.65 ± 0.90	1.87 ± 0.91
$Z+0$b-jets	1.16 ± 0.16	1.28 ± 0.22
$Z+1$b-jets	2.10 ± 0.46	1.71 ± 0.47
$Z+2$b-jets	1.17 ± 0.24	0.96 ± 0.25



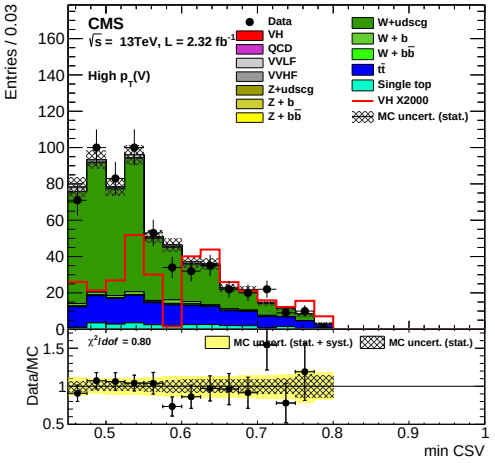
(A) $t\bar{t}$ control region



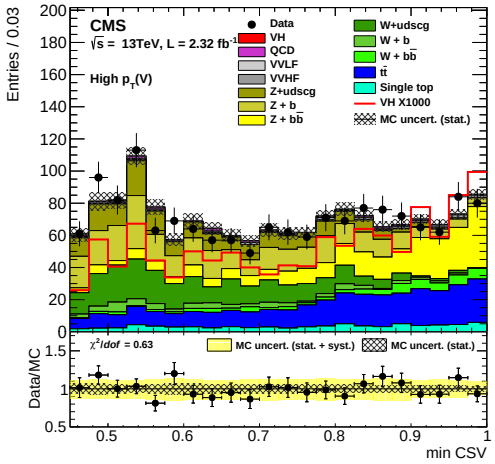
(B) multijet control region



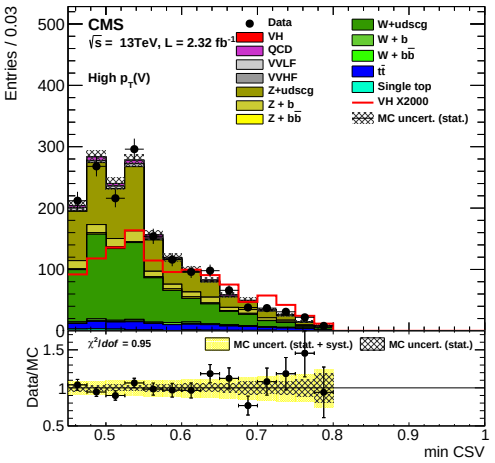
(C) W + b-jets control region



(D) W + light-jets control region

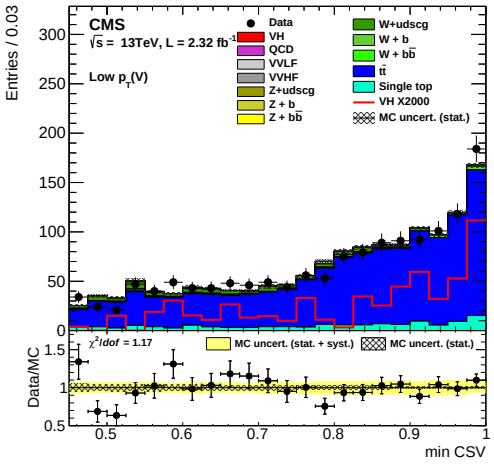


(E) Z + b-jets control region

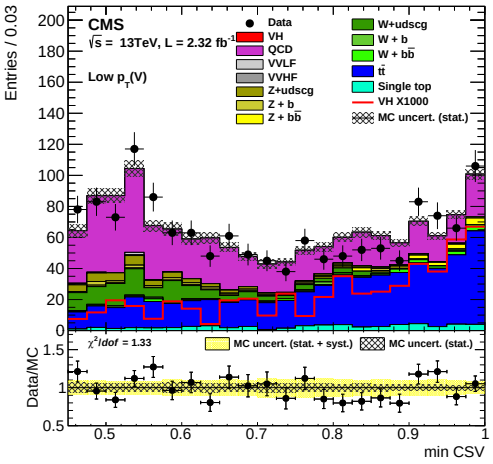


(F) Z + light-jets control region

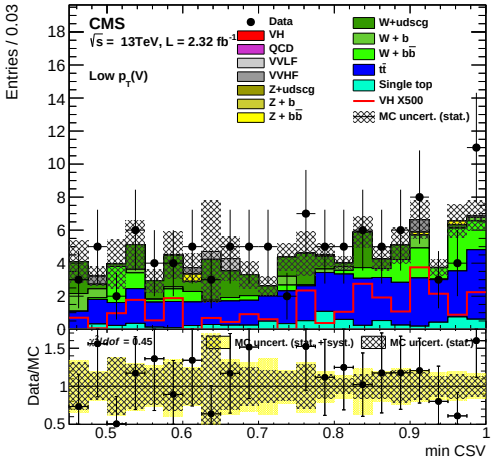
FIGURE 4.24: Post-fit $\mathrm{CSV}_{\mathrm{min}}$ distributions for the six control regions in the high $p_T(Z)$ bin.



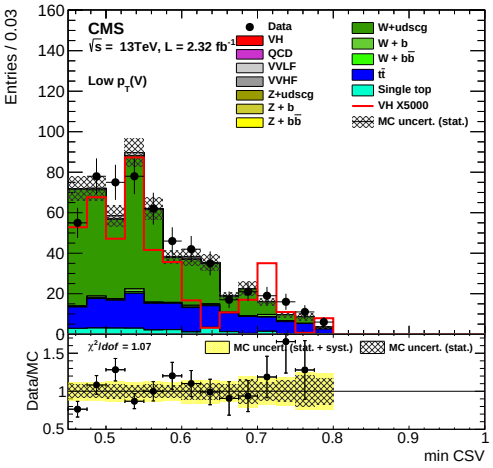
(A) $t\bar{t}$ control region



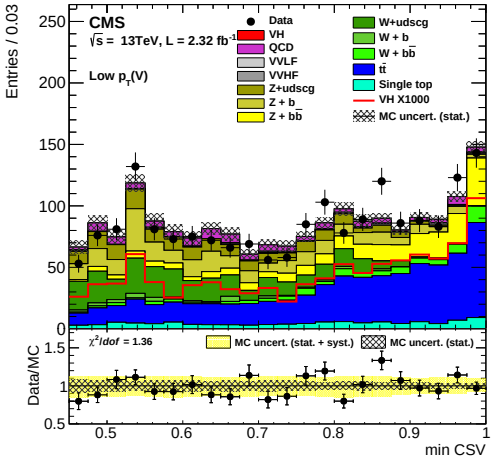
(B) multijet control region



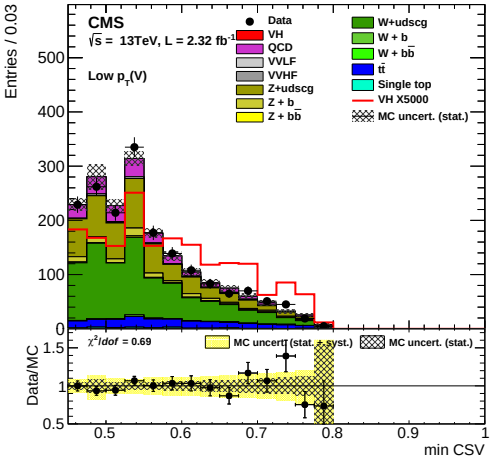
(C) W + b-jets control region



(D) W + light-jets control region

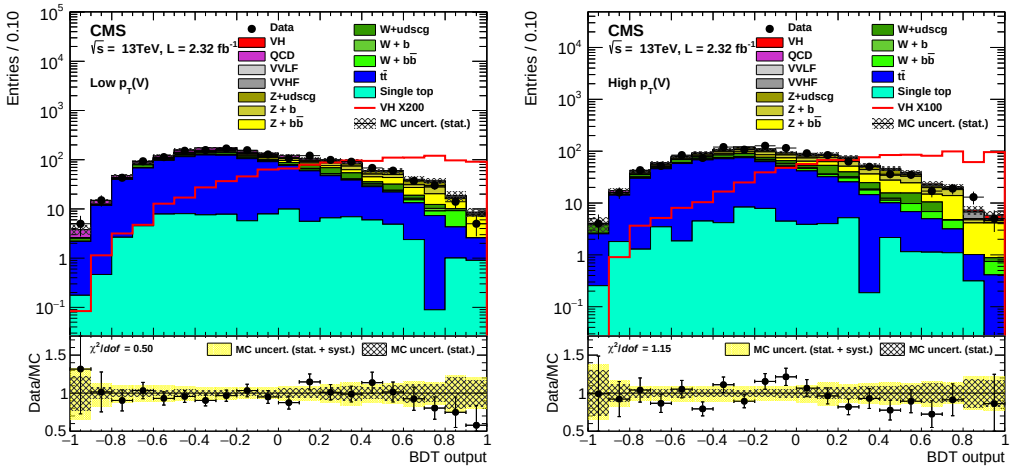


(E) Z + b-jets control region



(F) Z + light-jets control region

FIGURE 4.25: Post-fit $\mathrm{CSV}_{\mathrm{min}}$ distributions for the six control regions in the low $p_T(Z)$ bin.



(A) Low $p_T(Z)$ bin (B) High $p_T(Z)$ bin

FIGURE 4.26: Post-fit BDT output distributions for the low (left) and high (right) $p_T(Z)$ bin signal region.

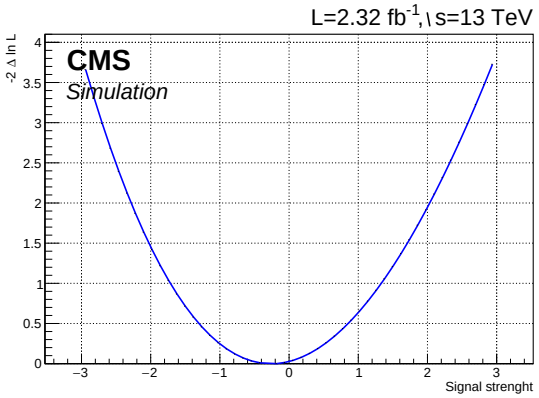


FIGURE 4.27: Likelihood scan of signal strength (μ) for $Z(\nu\nu)H(bb)$ with $2.32\,\text{fb}^{-1}$ of proton-proton collision at $\sqrt{s} = 13\,\text{TeV}$.

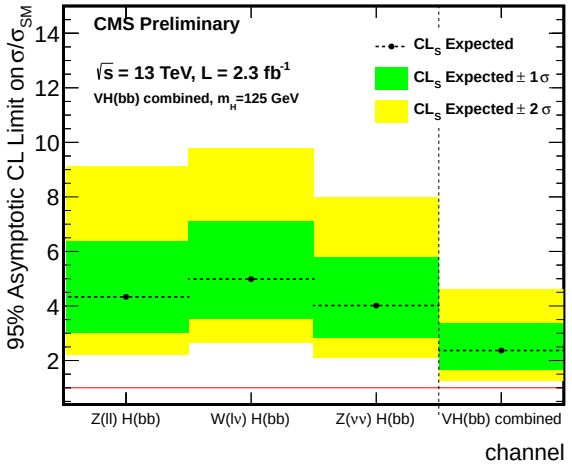


FIGURE 4.28: Expected 95% CL upper limit of the searches for $Z(\ell\ell)H(bb)$, $W(\ell\nu)H(bb)$, $Z(\nu\nu)H(bb)$, and their combination. The limit have been evaluated using $2.32\,\text{fb}^{-1}$ of proton-proton collision at $\sqrt{s} = 13\,\text{TeV}$.

4.11 Outlook for 2016

During the 2016, LHC is expected to deliver more than $20\,\text{fb}^{-1}$ of integrated luminosity and the trigger `HLT CaloMHTNoPU90 PFMET90 PFMHT90 IDTight BTagCSV0p72` is expected to be unprescaled during the whole data-taking. This will allow to keep the same acceptance (MET $> 120\,\text{GeV}$) in the future analysis based on the 2016 data-taking. For this reason, we can roughly estimate the sensitivity of the future analysis simply increasing the normalization of the simulation. Using $20\,\text{fb}^{-1}$, the expected upper limit is 1.37 times the SM Higgs boson cross-section. The expected significance is 1.39 standard deviations and the expected uncertainty on the signal strength is 0.68. Combining the searches for $Z(\ell\ell)H(bb)$, $W(\ell\nu)H(bb)$, $Z(\nu\nu)H(bb)$ with the Run-1 results [27], we expect to find the evidence of the $H \to bb$ decay with more than three standard deviations with the 2016 dataset.

Conclusion

A search for the $Z(\nu\nu)H(bb)$ associated production using $\int \mathcal{L} = 2.32\,\text{fb}^{-1}$ of proton-proton collisions at $\sqrt{s} = 13\,\text{TeV}$ with the CMS experiment has been presented. No excess have been observed and a 95% C.L. upper limit on the signal strength parameter $\mu = \sigma/\sigma_{SM}$ of 3.97 (3.95) is observed (expected). The fitted signal strength is $\mu = -0.25^{+1.47}_{-1.37}$. After the 2016 data-taking ($\sim 20\,\text{fb}^{-1}$), this search is expected to reach a sensitivity of about 1.37 standard deviations and to measure the parameter μ with an uncertainty of about 0.68. Combining this expected results with the Run-1 analysis [27] and with the $W(\ell\nu)H(bb)$ and $Z(\nu\nu)H(bb)$ channels, we expected to find the evidence of the $H \to bb$ decay with more than three standard deviations.

The acceptance of this and future analyses has been significantly enlarged by the trigger improvements that I developed during the Long Shutdown 1 of LHC. They include a fast tracking that can be run for every event accepted by the L1 trigger ($\sim 100\,\text{kHz}$). It has been used to tag pile-up jets in the early stage of the High Level Trigger. In $Z(\nu\nu)H(bb)$ trigger, it has allowed to improve the resolution on the missing transverse energy evaluated using calorimetric jets. The fast tracking has been also used to speed up the b tagging, allowing to run the online b tagging more often. These and other improvements allowed to increase the trigger efficiency of the fully hadronic $H(bb)$ analyses and will allow to keep a low MET threshold in the future high-luminosity $Z(\nu\nu)H(bb)$ analysis.

Appendices

Appendix A

The merged cluster splitter

A.1 Tracking efficiency inside a jet core

In the searches for the $H \to bb$, tracking plays a key role. Tracking is less performing within the cone of high p_T jets. Figure A.1 shows the offline b-tagging efficiency and fake rate, as a function of the jet p_T. Starting from $p_T \gtrsim 200\,\text{GeV}$, the b-tagging performance gets significantly worse. This happens because of a tracking inefficiency in the core of jets with $p_T \gtrsim 200\,\text{GeV}$. In that region, several tracks with high p_T and originating from the same primary vertex are concentrated within a narrow cone. The corresponding track hits are very close each other and this creates two problems:

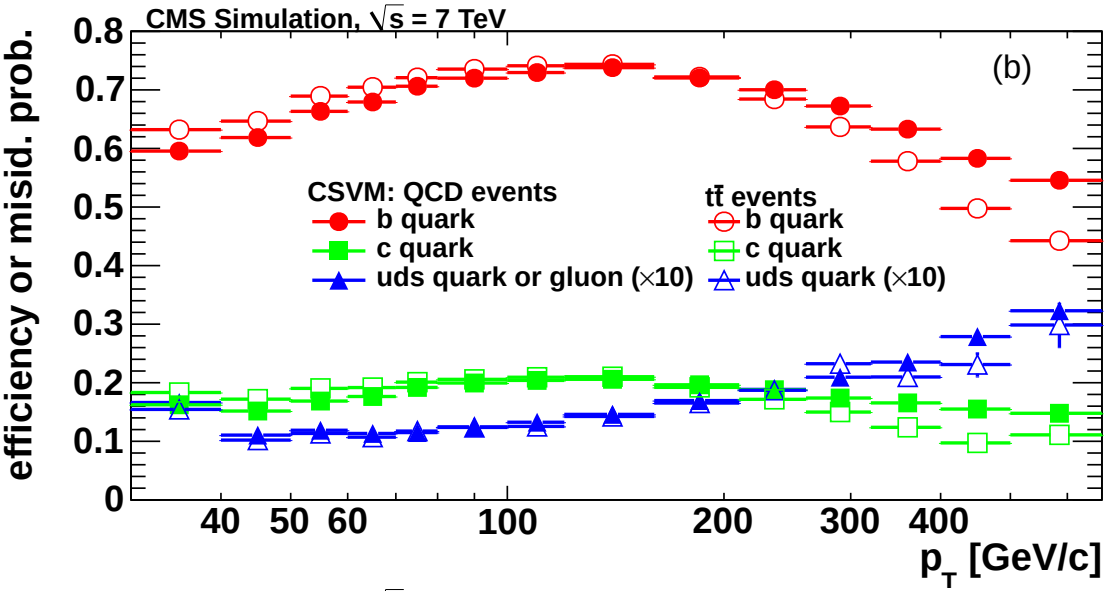


FIGURE A.1: Efficiency and misidentification probability for a b-tagging working point as a function of the jet p_T for b (red), c (green), and uds (blue) jets. The values indicated by the full (empty) markers have been obtained using a simulation of multijet ($t\bar{t}$) production with $\sqrt{s} = 7\,\text{GeV}$ [63].

- the tracking efficiency decreases because more fake tracks are reconstructed;
- the pixel hits can be merged in only one single hit.

The first problem has been solved by adding a dedicated step in the iterative tracking described in Sec. 2.3.1. The track seeds used in this iteration are pairs of pixel hits compatible with a track with $p_T > 10\,\text{GeV}$ originating from the primary and having $\Delta R(\text{track}, \text{jet}) < 0.2$ from jets with $p_T > 100\,\text{GeV}$. This iteration recovers the efficiency to reconstruct tracks with high p_T having a missing pixel hits and reduces the number of fake tracks reconstructed in the subsequent iterations.

The second issue has been solved with a novel algorithm that splits the merged pixel hits and is described in the next section.

A.2 The pixel cluster splitter algorithm

A.2.1 The problem

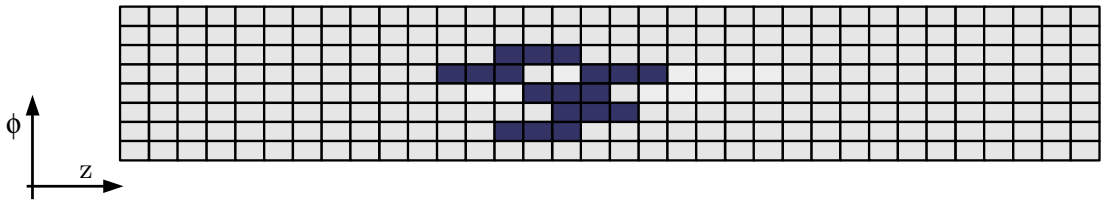
In the pixel detector, each track hit is localized by a pixel cluster, i.e. a set of adjacent pixels with signal, including both side-by-side and corner-by-corner pixels (see Sect. 2.3.1). When two or more tracks are very close, their pixel clusters may be reconstructed as a single hit that will be assigned to a single track. This degrades the tracking performance for two reasons.

- As the hit will be assigned to only one track, the remaining tracks will be reconstructed with a missing pixel hit, usually in the innermost layer, reducing both the quality and the impact parameter resolution of those tracks.
- The position of the single reconstructed hit will be wrong because it is evaluated merging different pixel clusters. This error may generate a fake impact parameter to the primary tracks, increasing the b-tagging fake-rate, and producing a wrong impact parameter for the secondary vertex tracks, degrading the b-tag efficiency.

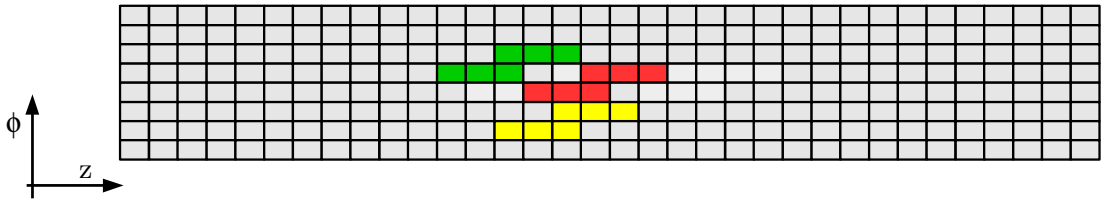
To solve this problem we need an algorithm that is able to split the merged pixel clusters, as shown in Fig. A.2.

A.2.2 A few considerations

Splitting merged clusters is a difficult problem because of the large number of degree of freedom. However we can exploit some constraints, like the number of tracks and the expected size of the clusters associated to each track along the two coordinates of the pixel modules, z and φ.



(A) A merged pixel cluster.



(B) A merged pixel cluster after the splitting.

FIGURE A.2: Three merged pixel cluster before (A) and after (B) the splitting. In Figure (A), the gray (blue) squares represent the pixels without (with) signal. In Figure (B), the green, red, and yellow squares represent the pixels associated with the three split clusters.

- **Cluster length.** Inside the jet core, all tracks are approximately parallel along η and therefore the expected size along the z-coordinate of all clusters is determinate by Eq. (3.1), already discussed in Sec. 3.2.1 (Fig.3.3 and 3.4):

$$\text{cluster length} = 1 + \frac{1.9}{\tan\theta}, \tag{A.1}$$

  where θ is the azimuthal angle of the jet and 1.9 is the thickness-to-width ratio of the pixel sensors.

- **Cluster width.** Tracks in the jet core have usually $p_T \gtrsim 1\,\text{GeV}$ and hence they hit the pixel layers almost orthogonally. For this reason, the expected cluster size along the φ coordinate is always of two pixels, because of the Lorentz angle effect (see Sect. 2.2.2).

- **Number of expected tracks.** The number of expected tracks in the merged cluster can be determinate from the total cluster charge observed. Since the incident angle of the tracks is approximately known, we can estimate the charge deposited by each track and hence obtain the number of tracks hitting the merged cluster.

Given these constraints, we can formulate the problem in the following way: we want to find the position of N tracks (x^{trk}, y^{trk}) on the pixel module, that minimize the χ^2 defined as

$$\chi^2 = \sum_{i \in pixels} \left(\rho_i^{meas} - \sum_{j \in tracks} \rho_i^{exp}(x_j^{trk}, y_j^{trk}) \right)^2, \tag{A.2}$$

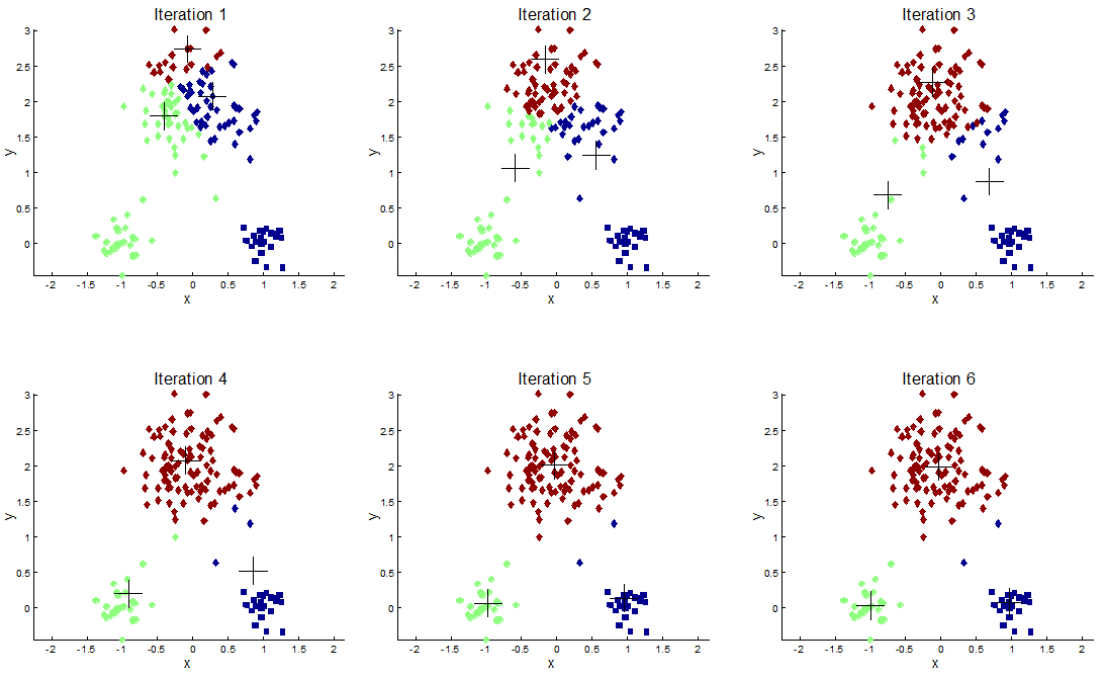


FIGURE A.3: Example of iterations of the k-means algorithm with $k = 3$. The small squares indicate the points $\vec{z_1}, \ldots, \vec{z_N}$ and the black crosses are for the cluster positions $\vec{x_1}, \vec{x_2}, \vec{x_3}$. The three colors (red, green, blue) indicate the association of the points $\vec{z_i}$ with the three clusters [90].

where ρ_i^{meas} is the charge measured in the pixel i, and $\rho_i^{exp}(x_j^{trk}, y_j^{trk})$ is the expected charge deposited in the pixel i by the track j, centered in (x_j^{trk}, y_j^{trk}).

This problem can be solved testing all the possible solution (x_j^{trk}, y_j^{trk}) but this is very slow because of the large number of degrees of freedom. For this reason, we solved it in a different way, based on the k-means algorithm.

A.2.3 The k-means algorithm

The k-means algorithm [89] is an iterative algorithm that, given a set of N points $\vec{z_1}, \ldots, \vec{z_N}$, localize the positions of k clusters $\vec{x_1}, \ldots, \vec{x_k}$ minimizing the distance:

$$d^2 = \sum_{i \in points} (\vec{z_i} - \vec{x_i})^2 \,, \tag{A.3}$$

where $\vec{x_i}$ is defined as the position of the cluster closer to the point $\vec{z_i}$. Figure A.3 is an example of the application of the k-means algorithm.

The k-means algorithm alternates two steps:

- associate each point $\vec{z_i}$ with the closest cluster $\vec{x_j}$;
- recompute the cluster positions $\vec{x_1}, \ldots, \vec{x_k}$ as the means of the points assigned to each cluster.

The algorithm starts with random values of $\vec{x_1}, \ldots, \vec{x_k}$ and ends when the cluster positions do not change between two iterations.

A.2.4 The pixel cluster splitter

The k-means algorithm has been modified to take into account some specific features of the cluster:

1. in the k-means algorithm, the distance d^2 is symmetric under rotations in the (z, φ) plane, but the shape of the expected charge distribution deposited by each track does not;
2. the number of points assigned to a cluster is totally free, but the total charge deposited by a track is approximately fixed and known;
3. the positions of the points are on a continuum, but the pixels have discrete positions and multiple tracks could deposit charge in the same pixel.

These problems have been solved in the following way:

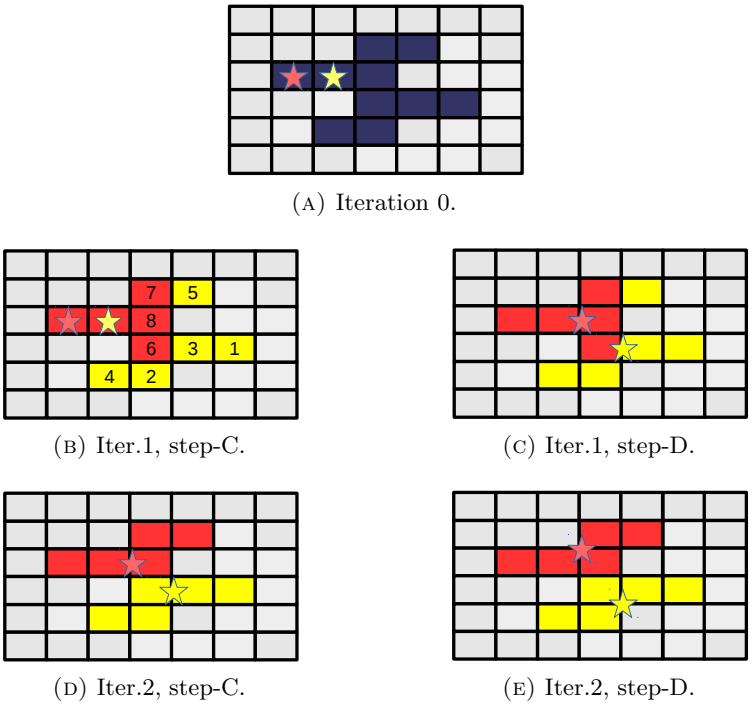
1. evaluate the distance d^2 using a metric that makes the expected shape of charge deposited by each track isotropic;
2. use a score to disadvantage the association of pixels with tracks that have already accumulated a large charge;
3. split each pixel into multiple points, having the same position, proportionally to the charge measured in that pixel.

In conclusion, after having evaluated the number of expected tracks in the merged cluster as the total cluster charge divided by the expected charge deposited by each track and having assigned random initial positions to the track positions, the pixel cluster splitter algorithm works iterating the following steps.

A. Compute the distance of each pixel-track pair, in a metric where the expected shape of the charge deposited by each track is isotropic.

B. Evaluate for each pair a score, taking into account both the distance and the charge accumulated by the corresponding track.

C. Find the farthest pixel from its second closest track and assign it to its closest track. This rule is useful to assign earlier the pixels that can be assigned to a track unambiguously. This step runs until all pixels are assigned.



(A) Iteration 0.

(B) Iter.1, step-C.

(C) Iter.1, step-D.

(D) Iter.2, step-C.

(E) Iter.2, step-D.

FIGURE A.4: Example of application of the pixel cluster splitter algorithm. The red and yellow stars correspond to the position $\vec{x_1}$ and $\vec{x_2}$ of the tracks, respectively. The red and yellow squares are the pixel assigned to the track 1 and 2, respectively. The blue squares are the unassigned pixels with signal, and the gray squares are the pixels without signal. Figure (A) shows the configuration before the application of the pixel cluster splitter. Figure (B) and (C) represent the assignment of the pixels to the tracks and the re-computation of the track positions in the first iteration, and Figure (D) and (E) are for the second iteration. The steps are described in the text. The numbers reported in Figure (B) correspond to the assignment order of the pixels to the tracks.

D. Update the track positions calculating the means of their assigned pixels.

The algorithm ends when the track positions are stabilized. Figure A.4 shows an example of the application of the pixel cluster splitter. In the example, two merged clusters are split in two iterations.

A.3 The results

The new step in the iterative tracking and the pixel cluster splitter have improved significantly the tracking performance inside the jet cores. Figure A.5 compares the tracking efficiency as a function of the ΔR between tracks and high-p_T jets both for the baseline and for the improved tracking using the pixel cluster splitter. The tracking efficiency at $\Delta R \sim 0.01$ has been improved from $\sim 75\%$ to $\sim 86\%$. The plot reports also that using the ideal cluster splitting (MC truth) the tracking efficiency would be $\sim 90\%$. It means that large part of the inefficiency related with the merged pixel clusters has been recovered and that the main part of the remaining inefficiency is not attributed to

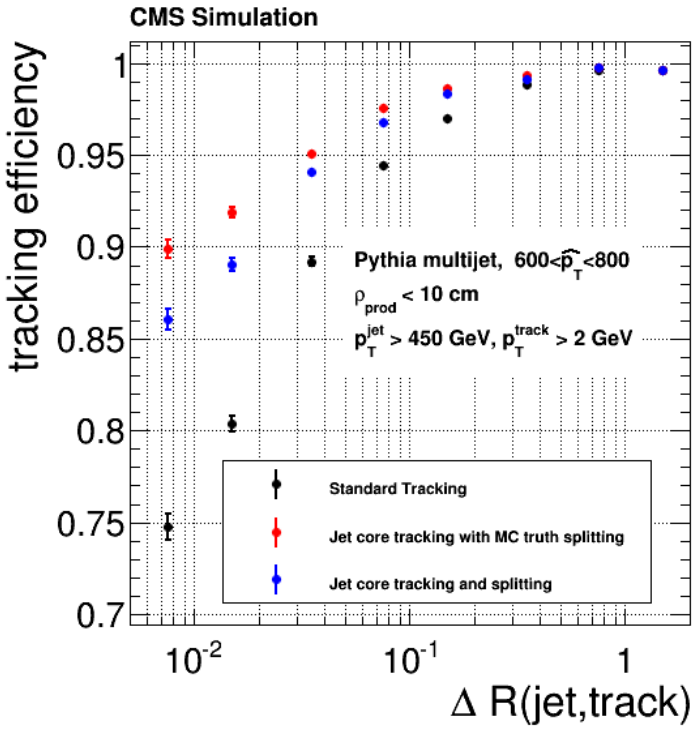


FIGURE A.5: Tracking efficiency as a function of ΔR between tracks and jets, using the baseline tracking (black), the improved tracking (blue), and the improved tracking with the MC-truth pixel-cluster splitting (red). The plot has been produced with a simulation of multijet production with 600 GeV $< \hat{p_T} <$ 800 GeV. Only jets with $p_T >$ 450 GeV and tracks with $p_T >$ 2 GeV and originating within 10 cm from the beam line have been considered.

the merged cluster issue but to the large combinatorial due to the high number of hits concentrated in a small amount of space.

Figure A.6 shows the tracking efficiency as a function of the distance between the beam line and the originating point of the track. For tracks produced with $\rho \sim 5$ mm the tracking efficiency has been increased from 85% to 93%. The increase of the tracking efficiency on the displaced tracks is important for improving the b-tagging performance.

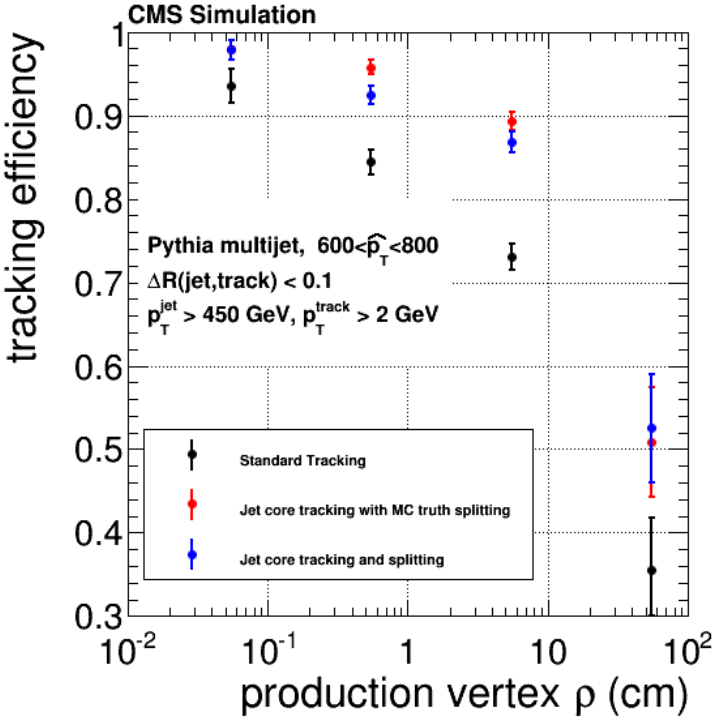


FIGURE A.6: Tracking efficiency as a function of the distance between the beam line and the originating point of the track, for the baseline tracking (black), the improved tracking (blue), and the improved tracking with the MC-truth pixel-cluster splitting (red). The plot has been produced with a simulation of multijet production with $600\,\text{GeV} < \hat{p_T} < 800\,\text{GeV}$. Only jets with $p_T > 450\,\text{GeV}$ and tracks with $p_T > 2\,\text{GeV}$ have been considered.

Appendix B

Evaluation of trigger rate using simulations with pile-up

B.1 Introduction

One of the difficulties in the preparation of the Run-2 HLT menu was the estimation of the HLT rate. As no 13-TeV data was available and LHC was expected to reach quickly $\mathcal{L} = 1.4 \cdot 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$, the trigger rate had to be estimated only using simulations. This Appendix describes three methods used to estimate the trigger rate using simulation. The third one is the novel method that I have developed.

The first method is the simplest one and it has been used to measure the rate of the L1 trigger (100 kHz). The second method is an intermediate step useful to understand the subsequent method. The third method is the new way of measuring the rate using simulations and it has been actually used during the preparation of the Run-2 HLT menu.

B.2 Three methods to evaluate the trigger rate

At LHC, the bunch crossing rate depends only on the number of bunches inside the accelerator ($N_{bunches} \lesssim 2800$) and the beam revolution time ($T_{rev} \sim 90\,\mu$s):

$$R_{ZeroBias} = \frac{N_{bunches}}{T_{rev}} \sim 30\,\mathrm{MHz}.$$

It is also called zero-bias rate because it would correspond to the rate of a trigger without any selection. The rate of a generic trigger T can be evaluated as:

$$R_T = P(T|\mu) \cdot R_{ZeroBias} \tag{B.1}$$

where $P(T|\mu)$ is the probability of the trigger T to accept an event, and μ is a parameter indicating the mean number of proton collisions per bunch crossing and is proportional to the instantaneous luminosity. The probability $P(T|\mu)$ can be split in:

$$P(T|\mu) = \sum_{n=1}^{\infty} P(T|\mu, n)P(n|\mu) \tag{B.2}$$

where $P(T|\mu, n)$ is the probability of the trigger T to pass an event having n collisions per bunch crossing, and $P(n|\mu)$ is the probability of having n collisions per bunch crossing and corresponds to the Poisson distribution with mean μ:

$$P(n|\mu) = \frac{\mu^n}{n!}e^{-\mu}. \tag{B.3}$$

Let us choose a variable to quantify the hardness of proton collisions. Any generator variable can be used for this purpose. For example, let consider the momentum of the leading scattered parton, $\hat{p_t}$, and let $f(\hat{p_t})$ be the probability density function of this variable. We can write explicitly the dependence of $P(T|\mu, n)$ from the $\hat{p_t}$ of each simultaneous proton collision:

$$P(T|\mu, n) = \int P(T|\mu, n, \hat{p_{t1}}, \ldots, \hat{p_{tn}}) \cdot f(\hat{p_{t1}}) \ldots f(\hat{p_{tn}}) \cdot d\hat{p_{t1}} \cdots d\hat{p_{tn}} \tag{B.4}$$

where $P(T|\mu, n, \hat{p_{t1}}, \ldots, \hat{p_{tn}})$ is the probability of the trigger to accept an event, given a set of $\hat{p_{ti}}$. The probability depends also on the number of simultaneous collisions n and on the mean number of collisions μ. The latter parameter is needed to take into account the out-of-time pile-up effects. Note that $f(\hat{p_t})$ is the same function for every collision and, therefore, the probability is symmetric under the exchange $\hat{p_{ti}} \leftrightarrow \hat{p_{tj}}$. Figure B.1 shows the distribution of $f(\hat{p_t})$ for a 13-TeV zero-bias simulation. Equation (B.4) is the common starting point of three methods that can be used to evaluate $P(T|\mu, n)$, and hence to evaluate the trigger rate.

B.2.1 First method

The simplest way to evaluate $P(T|\mu)$ is generating a large sample of zero-bias events with the number of simultaneous collisions distributed according to a Poisson distribution having mean μ. From this sample we can estimate:

$$P(T|\mu) \approx \frac{N(T\&\mu)}{N(\mu)}. \tag{B.5}$$

where $N(\mu)$ is the total number of events generated, and $N(T\&\mu)$ is the number of events that pass also the trigger T.

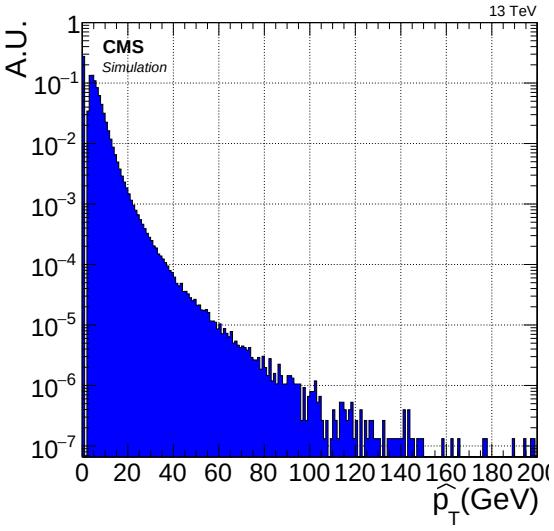


FIGURE B.1: Probability density function of $\hat{p_t}$ for a 13-TeV proton collision.

The limit of this method is that if you want to measure precisely ($< 10\%$) a small trigger rate ($r \sim 3\,\text{Hz}$) you need about one billion of zero-bias events. In addition, often, in the zero-bias samples no electroweak process are simulated, and this would underestimate the rate of leptonic triggers. For these reasons, this method is useful only to study triggers with large rate, like a L1 trigger ($r \sim 10\,\text{kHz}$), where the rate from QCD multijet event is dominant.

B.2.2 Second method

As shown in Fig. B.1, most of the zero-bias events are very soft and, therefore, many of the generated events do not pass the trigger. In the second method, instead, more events are generated in the high $\hat{p_t}$ region where the largest trigger rate contribution comes from, in an attempt to reduce the statistical uncertainty. In this case, simulations are produced generating one collision, binned in $\hat{p_t}$, and other $n-1$ unbinned collisions (pile-up). Using (B.4), we can express $P(T|\mu, n)$ as:

$$P(T|\mu,n) = \int P(T|\mu,n,\hat{p_{t1}},\hat{p_{t2}},\ldots,\hat{p_{tn}}) \cdot f(\hat{p_{t1}}) f(\hat{p_{t2}}) \ldots f(\hat{p_{tn}}) \cdot d\hat{p_{t1}} d\hat{p_{t2}} \cdots d\hat{p_{tn}} \tag{B.6a}$$

$$= \int \left[\int P(T|\mu,n,\hat{p_{t1}},\hat{p_{t2}},\ldots,\hat{p_{tn}}) f(\hat{p_{t2}}) \ldots f(\hat{p_{tn}}) \cdot d\hat{p_{t2}} \cdots d\hat{p_{tn}} \right] \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.6b}$$

$$= \int P(T|\mu,n,\hat{p_{t1}}) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.6c}$$

where $P(T|\mu, n, \hat{p_{t1}})$ is the probability that an event passes the trigger, when the "first" proton collision have $\hat{p_{t1}}$ and there are n collisions per bunch crossing.

Substituting $P(T|\mu, n)$ in (B.2), we obtain:

$$P(T|\mu) = \sum_{n=1}^{\infty} \left(\int P(T|\mu, n) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \right) \cdot P(n|\mu) \tag{B.7a}$$

$$= \int \left(\sum_{n=1}^{\infty} P(T|\mu, n) \cdot P(n|\mu) \right) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.7b}$$

$$= \int P(T|\mu) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.7c}$$

$$\tag{B.7d}$$

Let us suppose that the sample has been generated in m bins of $\hat{p}_t$. Then, the integral can be calculated as:

$$P(T|\mu) = \int P(T|\mu, \hat{p_{t1}}) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.8a}$$

$$= \sum_{j=1}^{m} \int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} P(T|\mu, \hat{p_{t1}}) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.8b}$$

$$= \sum_{j=1}^{m} \frac{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} P(T|\mu, \hat{p_{t1}}) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}}}{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} f(\hat{p_{t1}}) d\hat{p_{t1}}} \cdot \int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} f(\hat{p_{t1}}) d\hat{p_{t1}} \tag{B.8c}$$

$$= \sum_{j=1}^{m} P_j(T|\mu) \cdot w_j \tag{B.8d}$$

where $[p_{t,\min}^{(j)}, p_{t,\max}^{(j)}]$ is the range of the j-th $\hat{p}_t$ bin and $P_j(T|\mu)$ is the probability that an event of the j-th bin is accepted from the trigger.

The terms $P_j(T|\mu)$ in (B.8d) can be estimated with simulations evaluating the ratio of the number of events that passes the trigger to the total number of simulated events:

$$P_j(T|\mu) = \frac{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} P(T|\mu, \hat{p_{t1}}) \cdot f(\hat{p_{t1}}) d\hat{p_{t1}}}{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} f(\hat{p_{t1}}) d\hat{p_{t1}}} \approx \frac{N_j(T\&\mu)}{N_j(\mu)}. \tag{B.9}$$

The term w_j in (B.8d) is the weight of the j-bin:

$$w_j = \int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} f(\hat{p_t}) d\hat{p_t} = \frac{\sigma_j}{\sigma_{tot}} \tag{B.10}$$

where σ_{tot} and σ_j are, respectively, the total and j-th bin cross sections.

The limit of this method is that it still needs very large statistics to obtain small statistical uncertainty in rate estimates. To understand this, let us consider a trigger that

accepts any event containing, at least, one collision with $\hat{p_t} > p_{\hat{t}X}$. The events that contribute to the trigger rate are those having the "first" collision with $p_{\hat{t}1} > p_{\hat{t}X}$, summed to the events with $p_{\hat{t}1} < p_{\hat{t}X}$ and at least one additional collision with $p_{\hat{t}i} > p_{\hat{t}X}$ with $i \neq 1$. The first set of events ($p_{\hat{t}1} > p_{\hat{t}X}$) is contained in the high $p_{\hat{t}1}$ bins and then is simulated with large statistics. The second set ($p_{\hat{t}1} < p_{\hat{t}X}$), instead, is in the low $p_{\hat{t}1}$ bins and then has high statistical uncertainty. As all collisions are equivalent, the latter rate contribution is dominant and is $n-1$ times larger than the first one. For this reason, the statistical uncertainty reduction of this method is very limited. This statistical uncertainty problem will be definitely solved by the third method.

B.2.3 Third method

The third method is similar to the second one, but here we pick as collision-1, or leading collision, the collision with highest $\hat{p_t}$. In this way, by definition, we have:

$$p_{\hat{t}1} > p_{\hat{t}i}, \text{ for } i \neq 1.$$

Note that in this case the probability density function $f(\hat{p_t})$ of the leading and additional collisions are different, and they depend on the number of collisions n. The two functions will be called, respectively, $f_L(\hat{p_t}, n)$ and $f_{PU}(\hat{p_t}, n)$.

Let us write $P(T|\mu, n)$ as a function of $\hat{p_t}$ of the leading collision:

$$P(T|\mu,n) = \tag{B.11a}$$

$$= \int P(T|\mu, n, p_{\hat{t}1}, p_{\hat{t}2}, \ldots, p_{\hat{t}n}) \cdot f_L(p_{\hat{t}1}, n) f_{PU}(p_{\hat{t}2}, n) \ldots f_{PU}(p_{\hat{t}n}, n) \cdot dp_{\hat{t}1} dp_{\hat{t}2} \cdots dp_{\hat{t}n} \tag{B.11b}$$

$$= \int \left[\int P(T|\mu, n, p_{\hat{t}1}, p_{\hat{t}2}, \ldots, p_{\hat{t}n}) f_{PU}(p_{\hat{t}2}, n) \ldots f_{PU}(p_{\hat{t}n}, n) \cdot dp_{\hat{t}2} \cdots dp_{\hat{t}n} \right] \cdot f_L(p_{\hat{t}1}, n) dp_{\hat{t}1} \tag{B.11c}$$

$$= \int P(T|\mu, n, p_{\hat{t}1}, \text{sorted}) \cdot f_L(p_{\hat{t}1}, n) dp_{\hat{t}1} \tag{B.11d}$$

where $P(T|\mu, n, p_{\hat{t}1}, \text{sorted})$ is the probability that the trigger accepts an event having the leading collision with $p_{\hat{t}1}$, and $f_L(\hat{p_t}, n)$ is the probability density function of $\hat{p_t}$ for the leading collision. The function $f_L(p_{\hat{t}1}, n)$ can be obtained as the probability density function of a generic collision $f(p_{\hat{t}1})$, multiplied by $F^{n-1}(p_{\hat{t}1})$ that is probability that the other $n-1$ collisions have $\hat{p_t} < p_{\hat{t}1}$:

$$f_L(p_{\hat{t}1}, n) = n \cdot f(p_{\hat{t}1}) \cdot P(\text{sorted}|n, p_{\hat{t}1}) = n \cdot f(p_{\hat{t}1}) \cdot F^{n-1}(p_{\hat{t}1}) \tag{B.12a}$$

$$\text{with } F(p_{\hat{t}1}) = \int_0^{p_{\hat{t}1}} f(\hat{p_t}) d\hat{p_t}. \tag{B.12b}$$

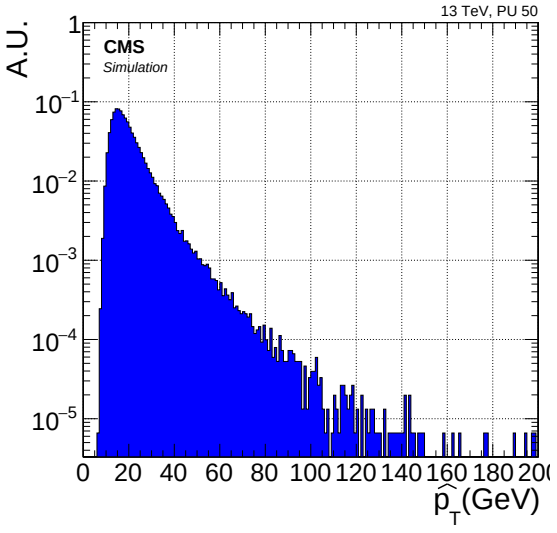


FIGURE B.2: Probability density function of the maximum $\hat{p_t}$ among fifty 13-TeV proton collisions. The events with $\hat{p_t} = 0\,\text{GeV}$ are those that cannot be modeled as a free collision, e.g. diffractive or semi-diffractive scattering.

where n is the normalization factor needed to get $f_L(\hat{p_{t1}})$ properly normalized. Basically, the factor takes into account that the probability of a random collision to be the leading one is $1/n$. Section B.3 demonstrates formally that $\int_0^\infty f_L(\hat{p_{t1}}, n) = 1$ and hence that n is the proper normalization. Figure B.2 shows the distribution of $f_L(\hat{p_t}, \mu)$ in a 13-TeV zero-bias sample with fifty simultaneous collisions.

Substituting Eq. (B.12) in (B.11), we obtain:

$$P(T|\mu, n) = \int P(T|\mu, n, \hat{p_t}, \text{sorted}) \cdot n \cdot f(\hat{p_t}) \cdot P(\text{sorted}|n, \hat{p_t}) d\hat{p_t} \tag{B.13a}$$

$$= n \int P(T|\mu, n, \hat{p_t}, \text{sorted}) \cdot P(\text{sorted}|n, \hat{p_t}) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.13b}$$

$$= n \int P(T, \text{sorted}|\mu, n, \hat{p_t}) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.13c}$$

where it has been applied the probability rule: $P(A|B,C)P(B|C) = P(A,B|C)$.

Substituting $P(T|\mu, n)$ in (B.2), we get:

$$P(T|\mu) = \sum_{n=1}^{\infty} \left(n \int P(T, \text{sorted}|n, \hat{p_t}) \cdot f(\hat{p_t}) d\hat{p_t} \right) P(n|\mu) \tag{B.14a}$$

$$= \int \left(\sum_{n=1}^{\infty} n \cdot P(T, \text{sorted}|n, \hat{p_t}) \cdot P(n|\mu) \right) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.14b}$$

$$= \int \left(\sum_{n=1}^{\infty} n \cdot P(T, \text{sorted}|n, \hat{p_t}) \cdot \frac{\mu}{n} \cdot \frac{\mu^{(n-1)}}{(n-1)!} e^{-\mu} \right) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.14c}$$

$$= \mu \int \left(\sum_{n=1}^{\infty} P(T, \text{sorted}|n, \hat{p_t}) \cdot P(n-1|\mu) \right) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.14d}$$

$$= \mu \int \left(\sum_{m=0}^{\infty} P(T, \text{sorted}|m+1, \hat{p_t}) \cdot P(m|\mu) \right) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.14e}$$

$$= \mu \int P^*(T, \text{sorted}|\mu, \hat{p_t}) \cdot f(\hat{p_t}) d\hat{p_t} \tag{B.14f}$$

where we have defined $P^*(T, \text{sorted}|\mu, \hat{p_t}) = \sum_{m=0}^{\infty} P(T, \text{sorted}|m+1, \hat{p_t}) \cdot P(m|\mu)$. The probability P^* is evaluated assuming that the number of collisions is generated with a Poisson distribution shifted by one. In other words, in the sample used to evaluate P^* the number of pile-up collisions has to be generated as a Poisson distribution.

In case of $\hat{p_t}$-binned samples, likewise to (B.8), $P^*(T|\mu)$ can be expressed as:

$$P^*(T|\mu) = \sum_{j=1}^{m} P_j^*(T, \text{sorted}|\mu) \cdot w_j \tag{B.15}$$

where the term w_j is the weight of the j-bin (see (B.10)) and, using (B.14f), $P_j^*(T|\mu)$ is estimated as:

$$P_j^*(T|\mu) = \frac{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} \mu \cdot P^*(T, \text{sorted}|\mu, \hat{p_t}) \cdot f(\hat{p_t}) d\hat{p_t}}{\int_{p_{t,\min}^{(j)}}^{p_{t,\max}^{(j)}} f(\hat{p_t}) d\hat{p_t}} \approx \mu \frac{N_j^*(T \& \text{sorted} \& \mu)}{N_j^*(\mu)}. \tag{B.16}$$

The trigger probability, and hence the rate, can be estimated with simulations evaluating the ratio of events that are both accepted by the trigger and "sorted", i.e. with $\hat{p_{t1}} > \hat{p_{ti}}$ for any $i \neq 1$. Technically, it can be easily achieved adding a fake filter in the trigger that rejects events containing a pile-up collision with $\hat{p_{ti}} > \hat{p_{t1}}$.

Note that, differently from the second method, in the case of a trigger that accept all events containing at least one collision with $\hat{p_{t1}} > \hat{p_{tX}}$, the trigger rate is totally determined by the high $\hat{p_t}$ bins that can be simulated with large statistics. In short, this method avoids simulating a huge number of soft events because usually they do not contribute to the trigger rate. The third method exploits better the $\hat{p_t}$-binned simulation

reducing substantially the statistical uncertainty and it has been actually used during the preparation of the Run-2 HLT menu of the CMS experiment.

B.3 Normalization of $f_L(\hat{p_t}, n)$

This section demonstrates that $\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} = 1$.

$$\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} = \tag{B.17a}$$

$$\int_0^\infty n \cdot f(\hat{p_t}) \cdot F^{n-1}(\hat{p_t})d\hat{p_t} = \tag{B.17b}$$

$$n\int_0^\infty f(\hat{p_t}) \cdot F^{n-1}(\hat{p_t})d\hat{p_t} = \tag{B.17c}$$

$$n\left[\left(F(\hat{p_t}) \cdot F^{n-1}(\hat{p_t})\right)|_0^\infty - \int_0^\infty F(\hat{p_t}) \cdot (n-1)F^{n-2}(\hat{p_t})f(\hat{p_t})d\hat{p_t}\right] = \tag{B.17d}$$

$$n\left[1 - \int_0^\infty (n-1)F^{n-1}(\hat{p_t})f(\hat{p_t})d\hat{p_t}\right] = \tag{B.17e}$$

$$n - (n-1)\int_0^\infty nF^{n-1}(\hat{p_t})f(\hat{p_t})d\hat{p_t} = \tag{B.17f}$$

$$n - (n-1)\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} \tag{B.17g}$$

From the last equation we get the normalization:

$$\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} = n - (n-1)\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} \tag{B.18a}$$

$$n\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} = n \tag{B.18b}$$

$$\int_0^\infty f_L(\hat{p_t}, n)d\hat{p_t} = 1 \tag{B.18c}$$

Appendix C

Additional plots of the search for the $Z(\nu\nu)H(bb)$

This Appendix shows some additional plots obtained for the search for the $Z(\nu\nu)H(bb)$ (Chap. 4). The plots are the distributions of the main discriminating variables, including the BDT output, in the signal and control regions. The plots include the preliminary scale factors shown in Sect. 4.7.3.

In the low $p_T(Z)$ bin, the plots have been obtained in:

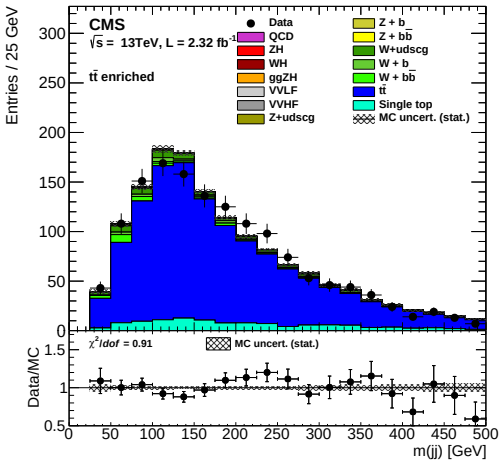
- $t\bar{t}$ control region (Fig. C.1),
- W + light-jets control region (Fig. C.2),
- $W + b$-jets control region (Fig. C.3),
- Z + light-jets control region (Fig. C.4),
- $Z + b$-jets control region (Fig. C.5),
- multijet control region (already shown in Fig. 4.23).

In the high $p_T(Z)$ bin, the plots have been obtained in:

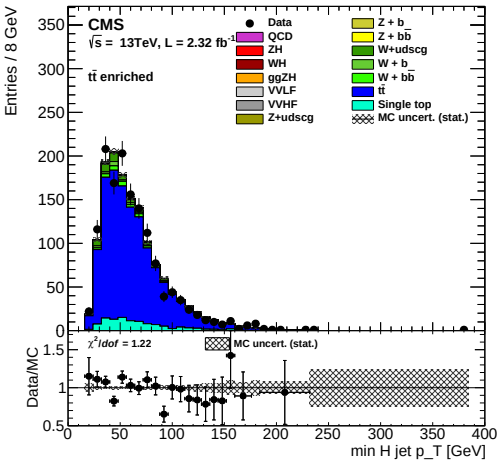
- $t\bar{t}$ control region (Fig. C.6),
- W + light-jets control region (Fig. C.7),
- $W + b$-jets control region (Fig. C.8),
- Z + light-jets control region (Fig. C.9),
- $Z + b$-jets control region (already shown in Fig. 4.22),

- multijet control region (Fig. C.10).

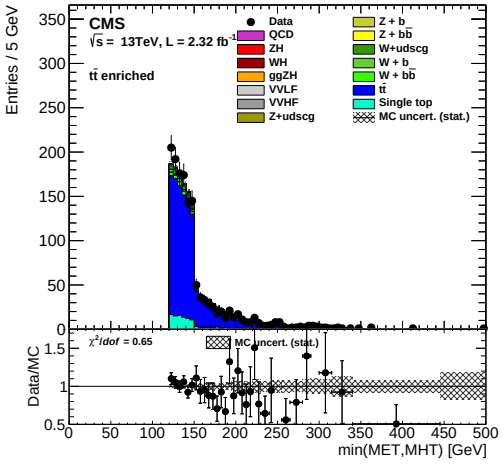
The distributions obtained in the signal regions are shown in Fig. C.11 (low $p_T(Z)$) and C.12 (high $p_T(Z)$).



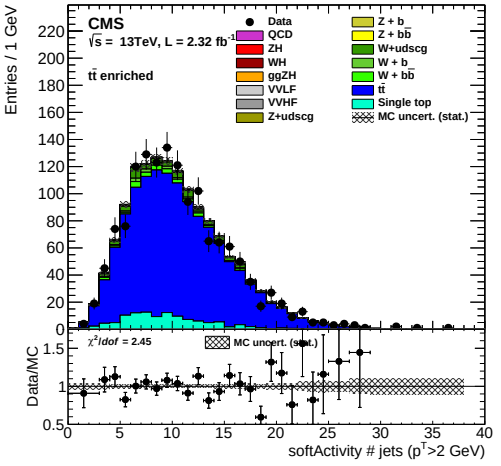
(A) m_{jj}



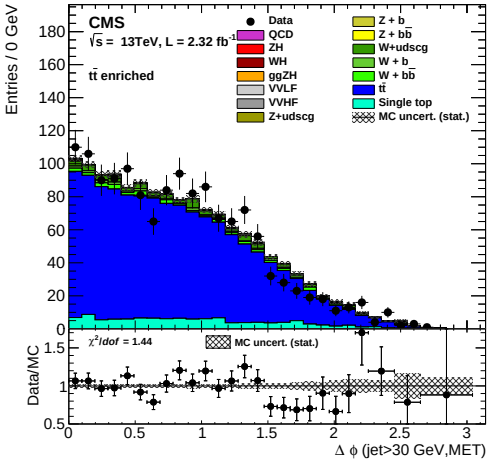
(B) min. Higgs jet p_T



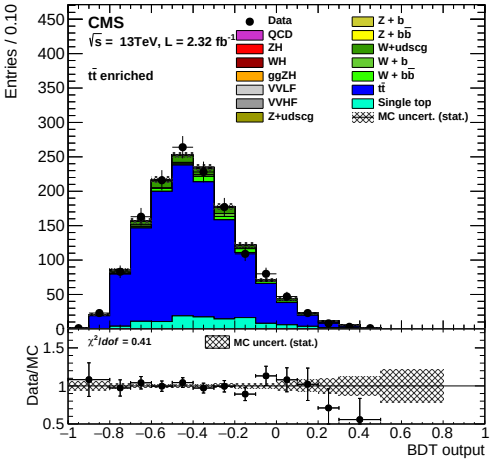
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

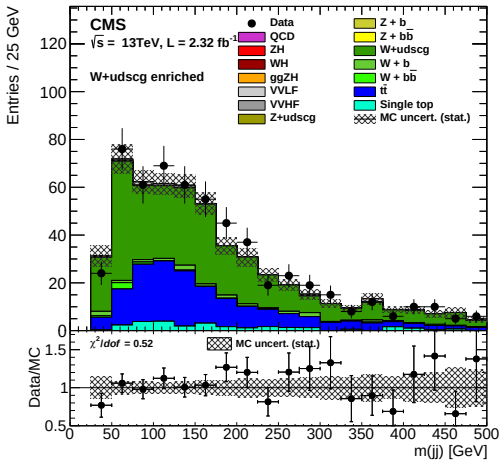


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

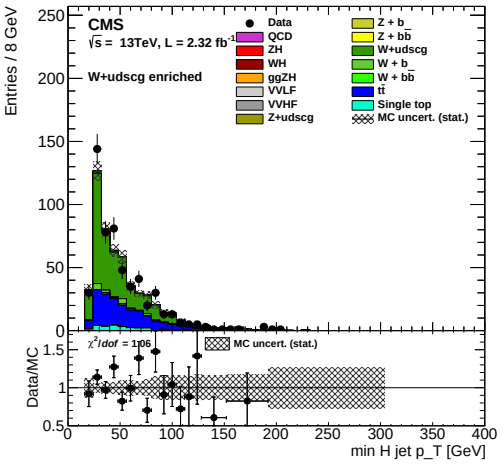


(F) BDT output

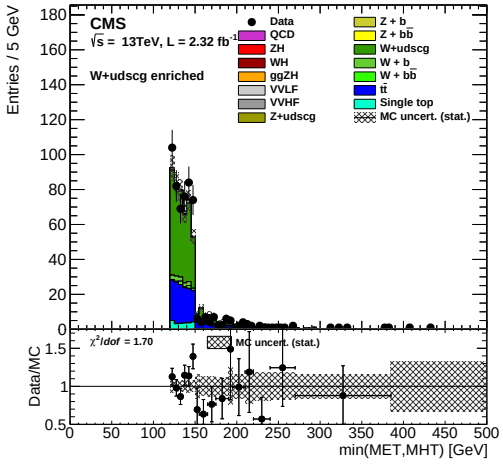
FIGURE C.1: Distributions of some variables for the $t\bar{t}$-enriched control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



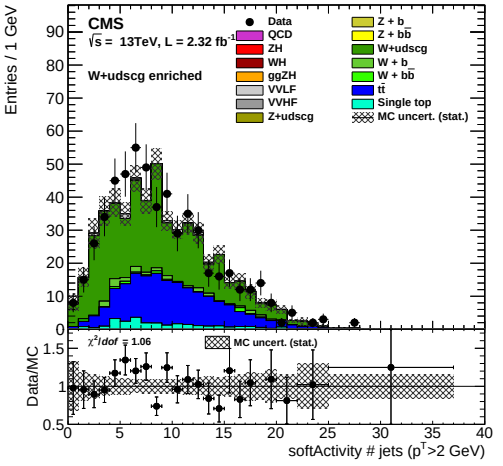
(A) m_{jj}



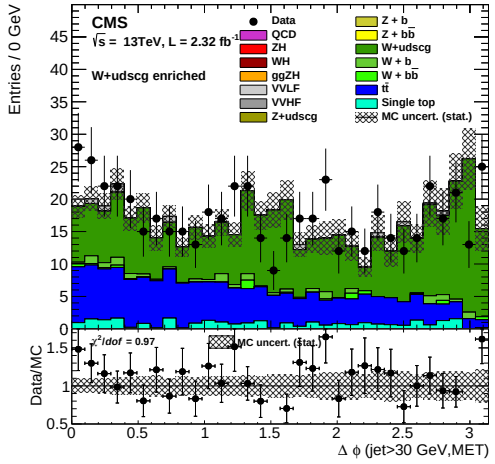
(B) min. Higgs jet p_T



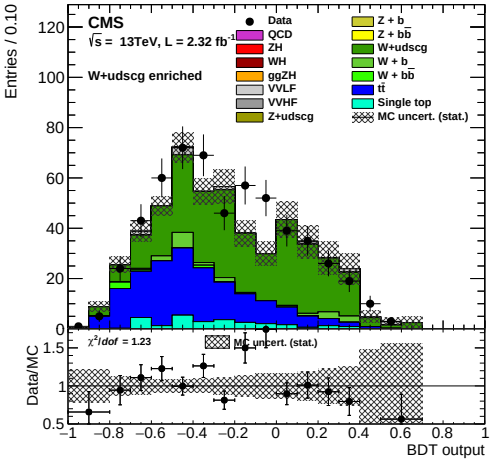
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

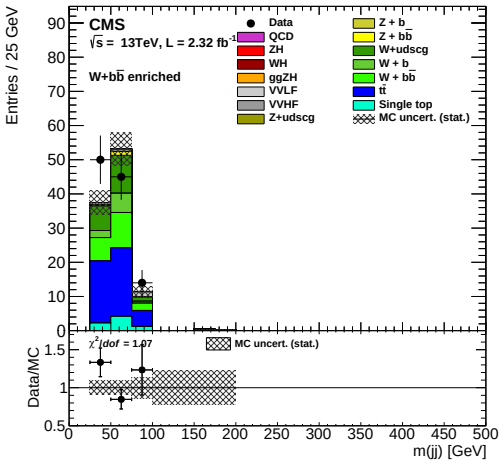


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

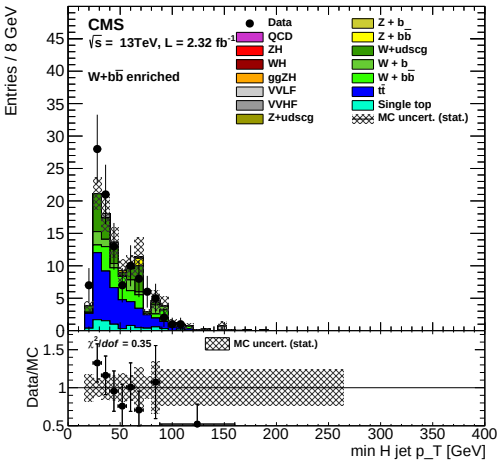


(F) BDT output

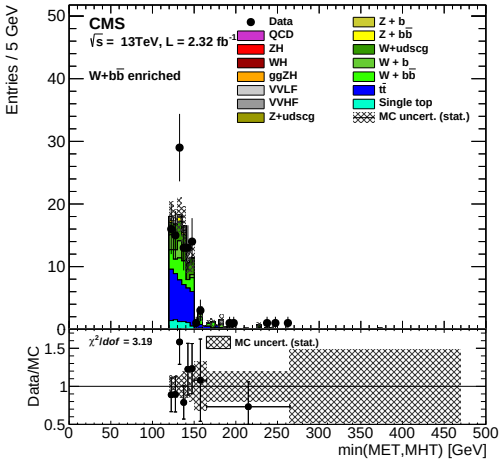
FIGURE C.2: Distributions of some variables for the W+light-jets control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



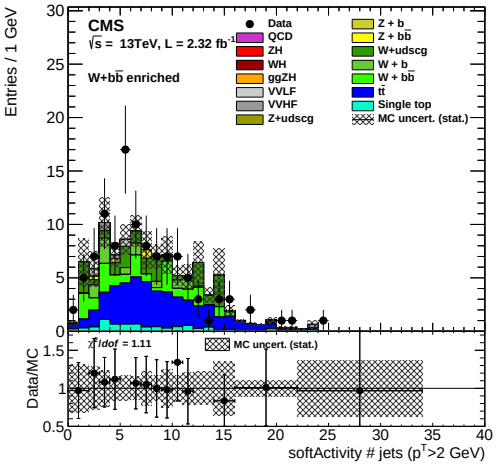
(A) m_{jj}



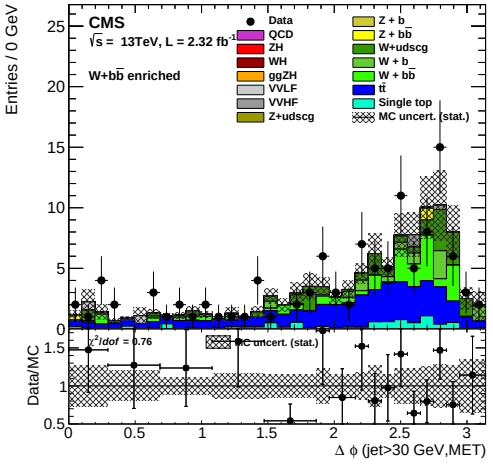
(B) min. Higgs jet p_T



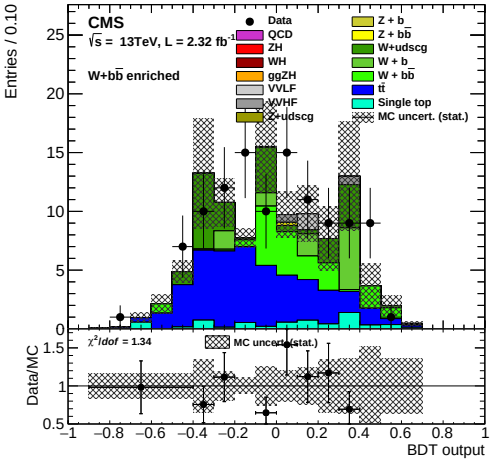
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

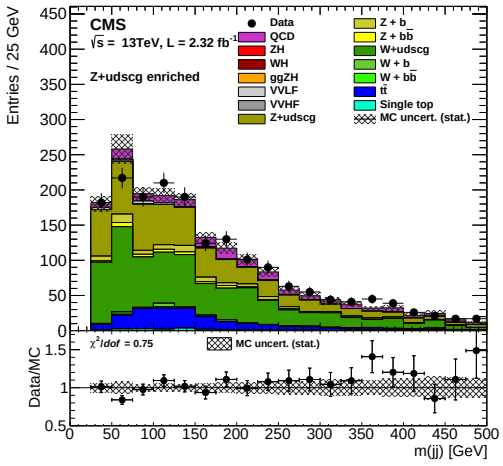


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

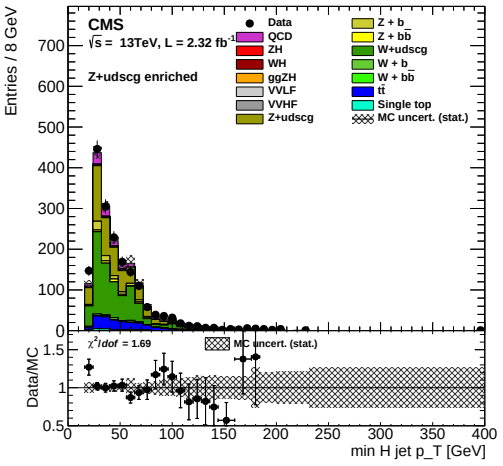


(F) BDT output

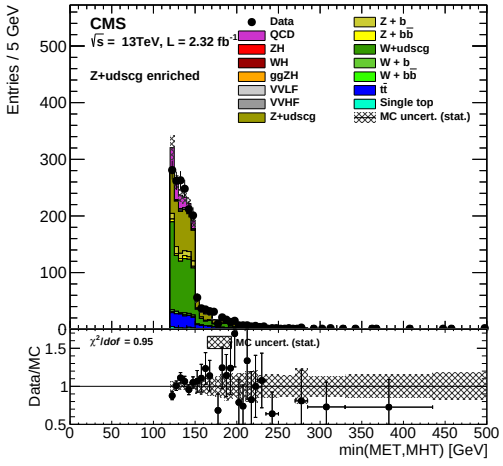
FIGURE C.3: Distributions of some variables for the W+b-jets control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



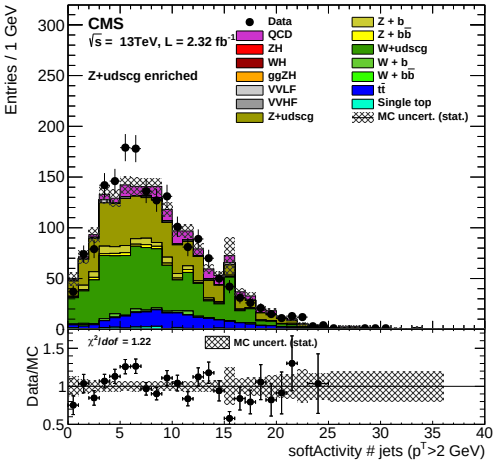
(A) m_{jj}



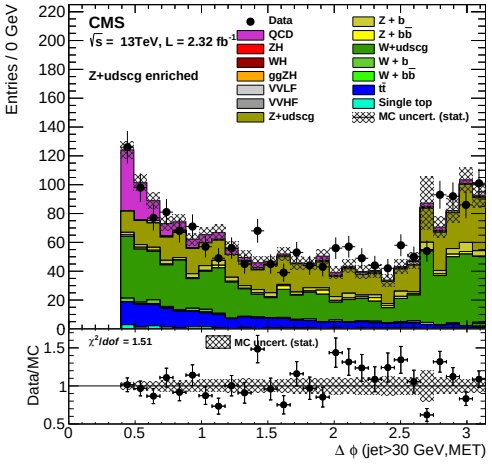
(B) min. Higgs jet p_T



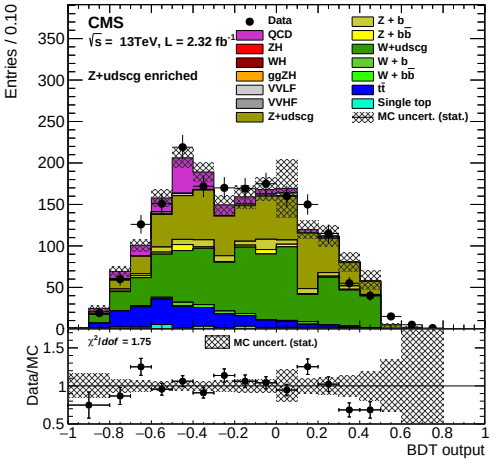
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

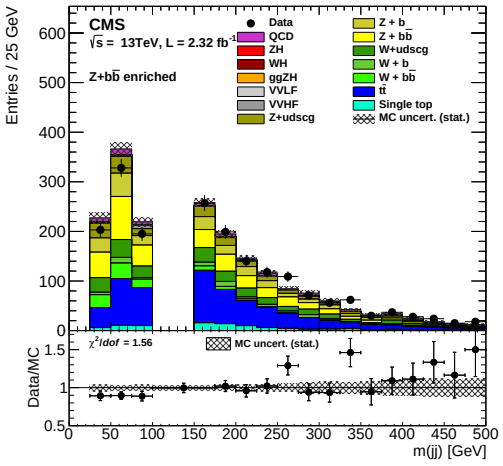


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

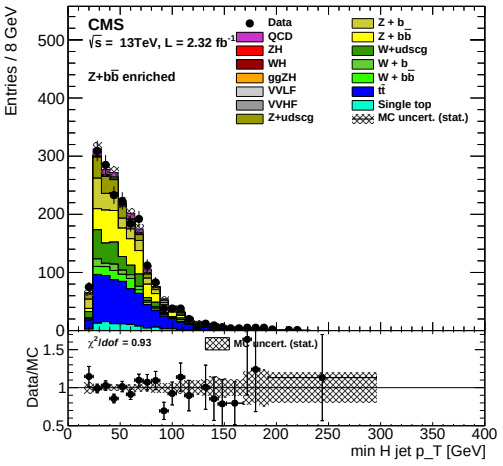


(F) BDT output

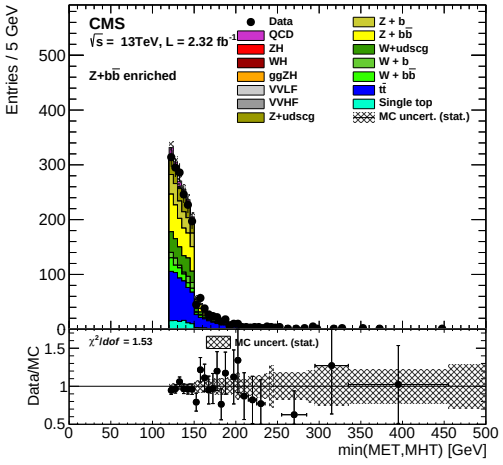
FIGURE C.4: Distributions of some variables for the Z+light-jets control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



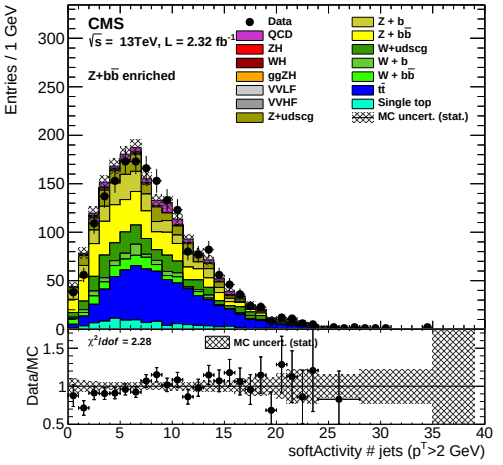
(A) m_{jj}



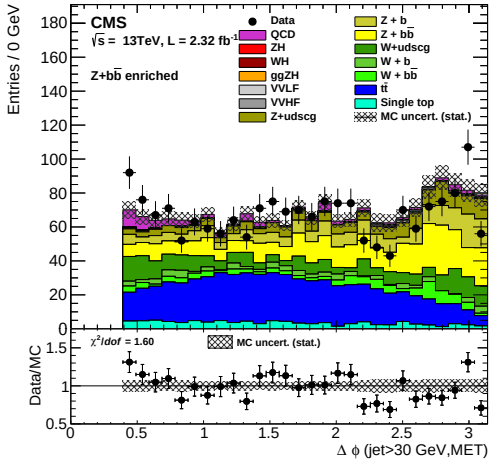
(B) min. Higgs jet p_T



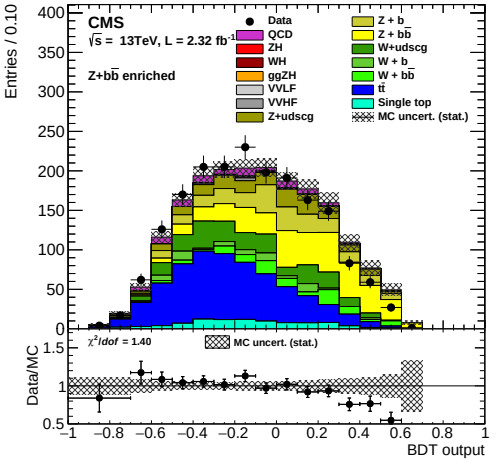
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

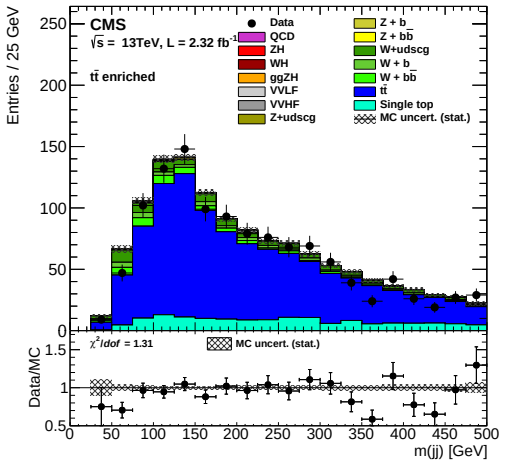


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

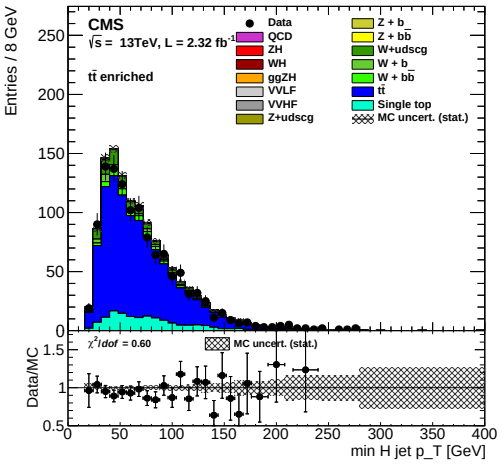


(F) BDT output

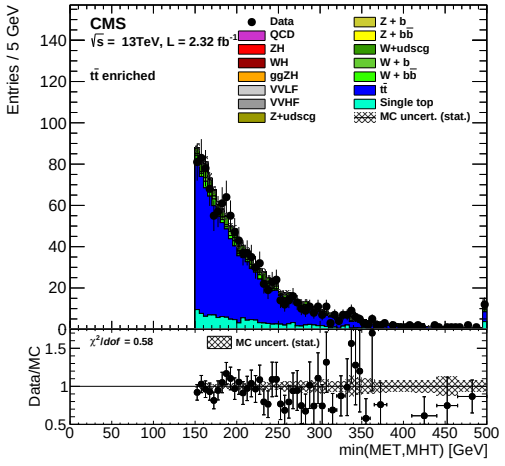
FIGURE C.5: Distributions of some variables for the Z+b-jets control region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



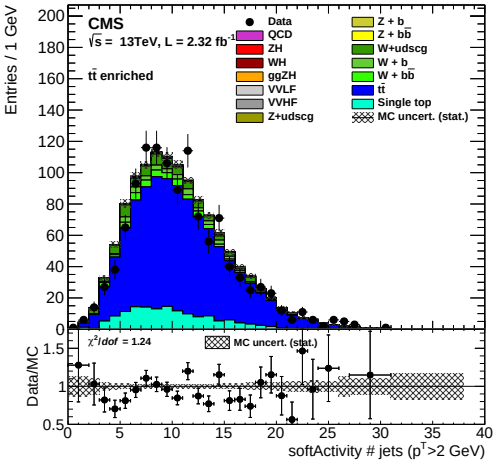
(A) m_{jj}



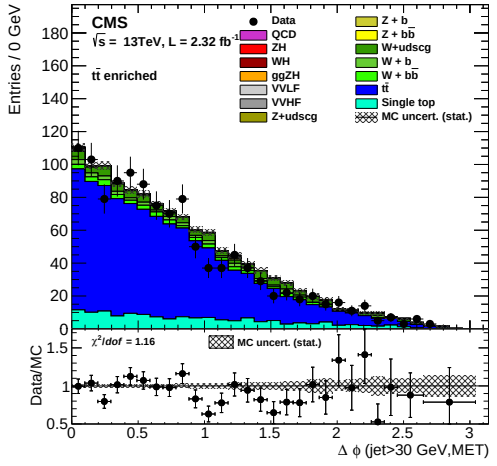
(B) min. Higgs jet p_T



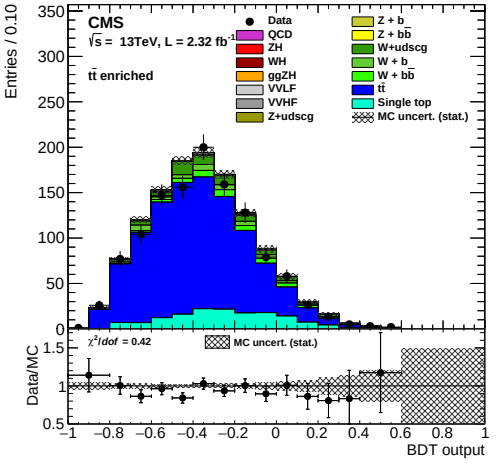
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

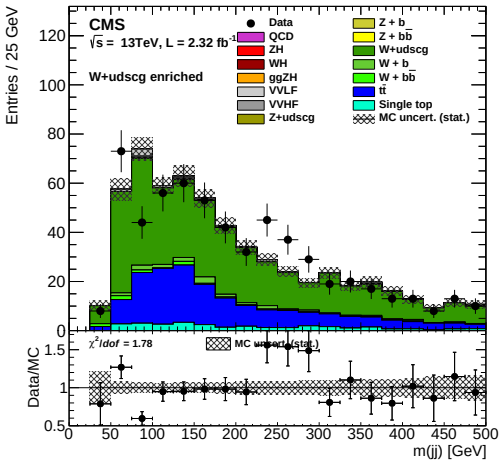


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

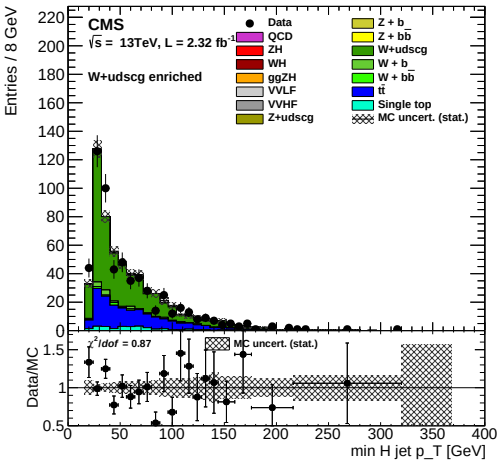


(F) BDT output

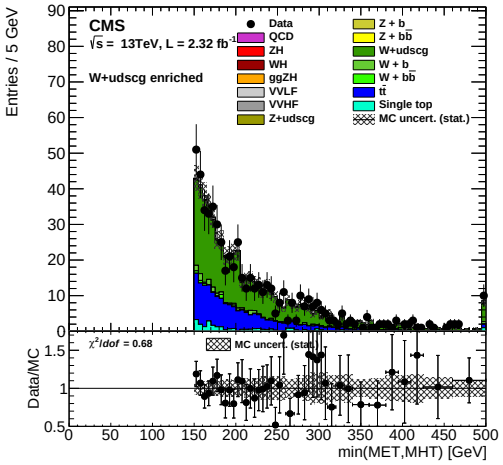
FIGURE C.6: Distributions of some variables for the $t\bar{t}$-enriched control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



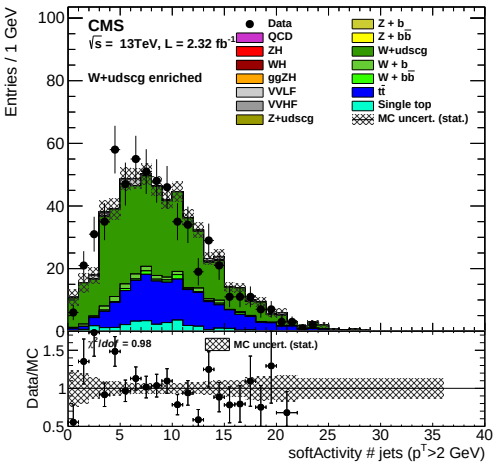
(A) m_{jj}



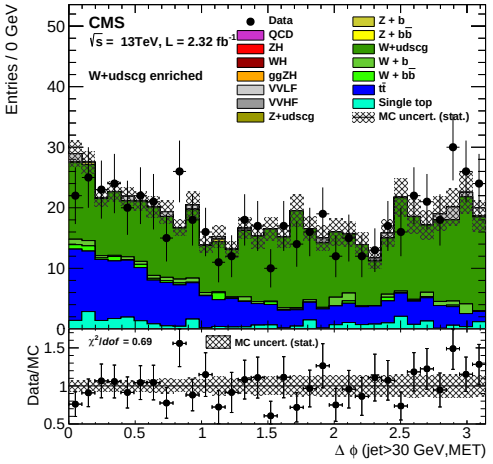
(B) min. Higgs jet p_T



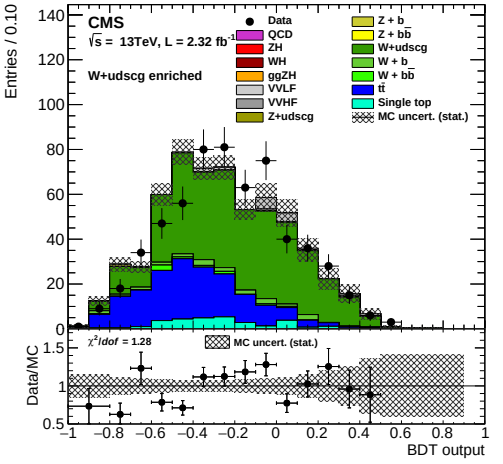
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

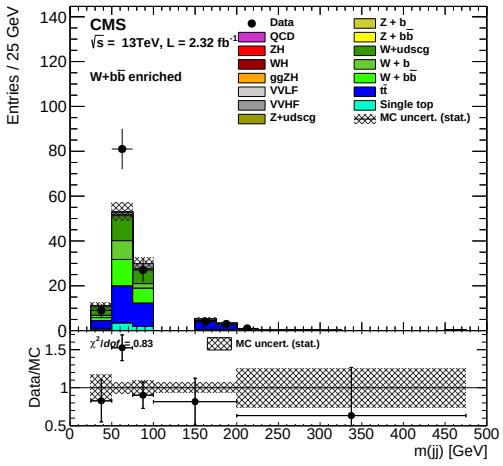


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

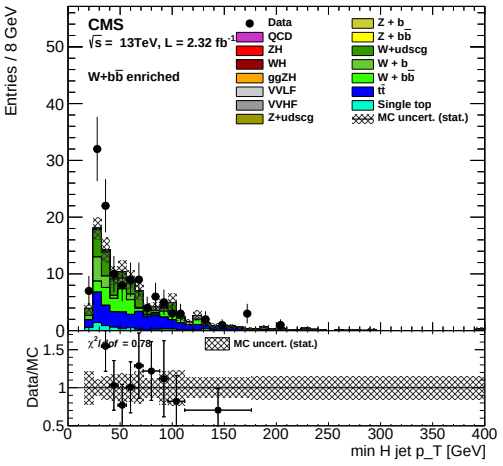


(F) BDT output

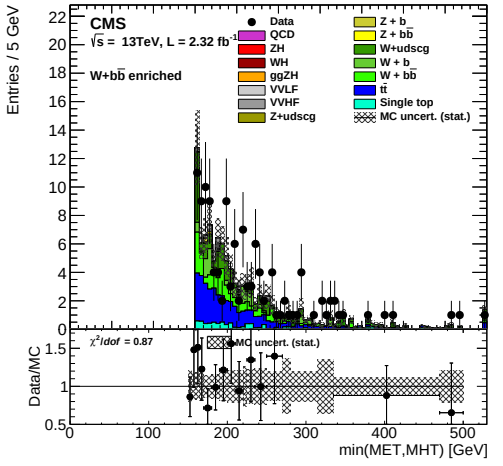
FIGURE C.7: Distributions of some variables for the W+light-jets control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



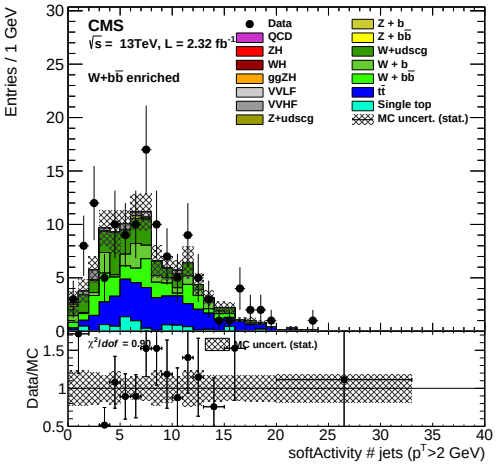
(A) m_{jj}



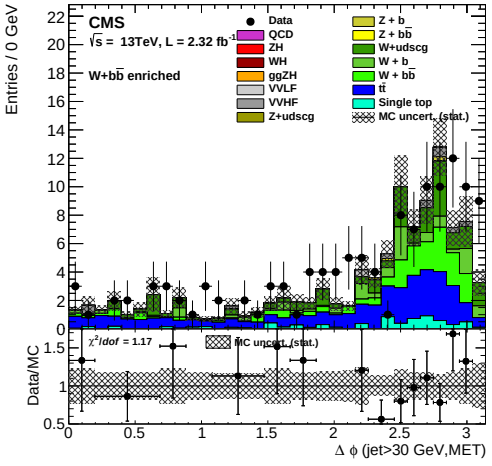
(B) min. Higgs jet p_T



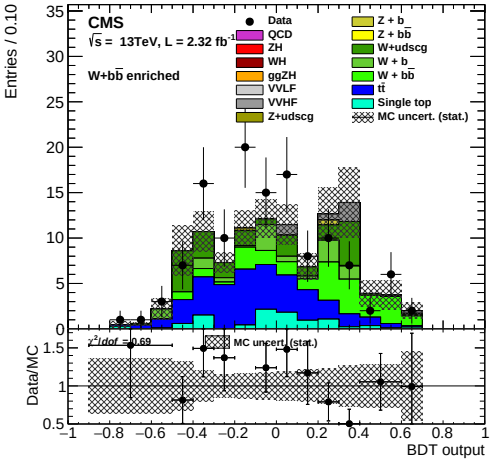
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

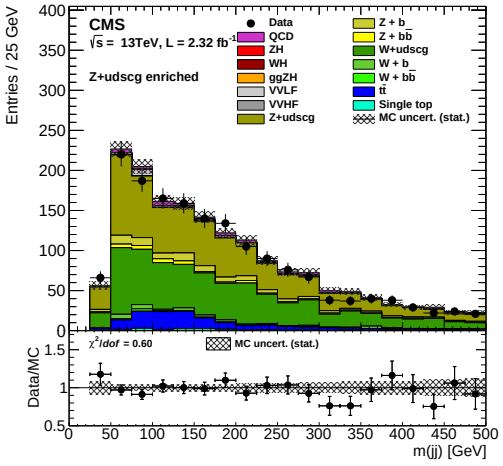


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

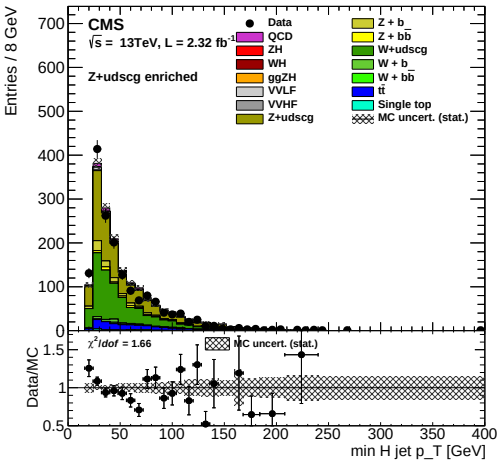


(F) BDT output

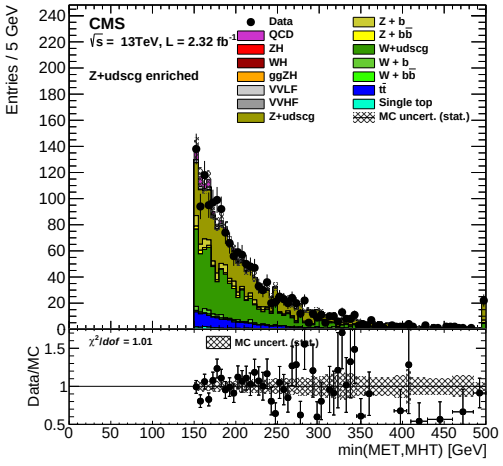
FIGURE C.8: Distributions of some variables for the W+b-jets control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



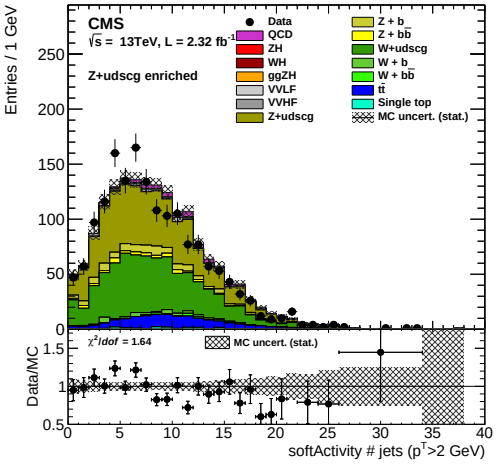
(A) m_{jj}



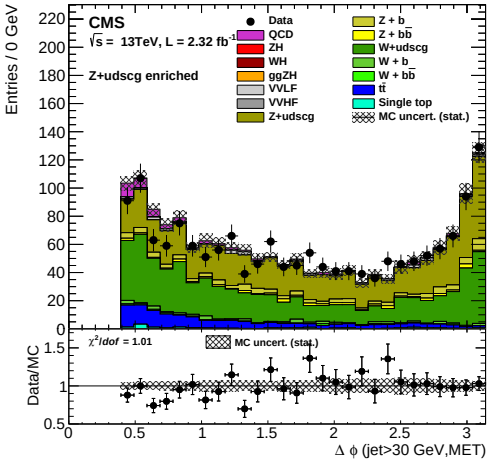
(B) min. Higgs jet p_T



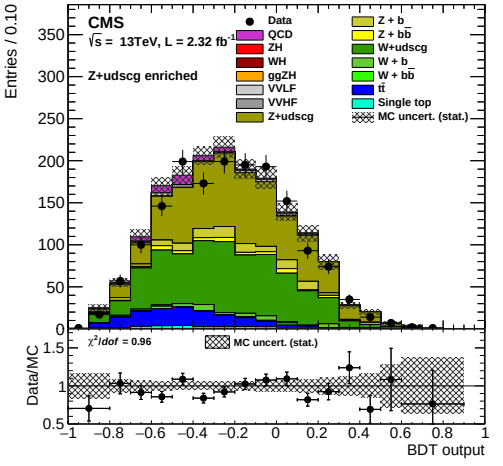
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

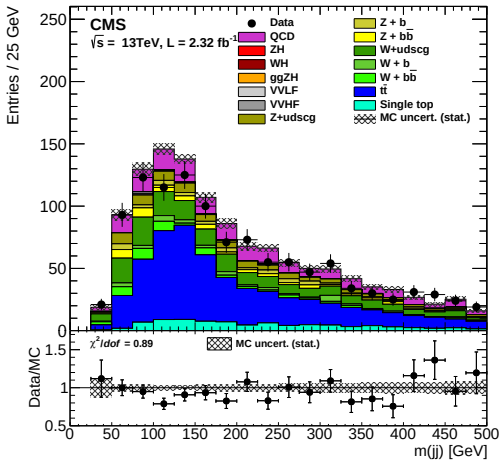


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

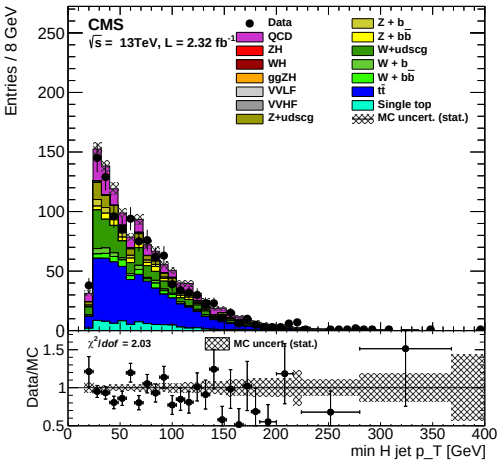


(F) BDT output

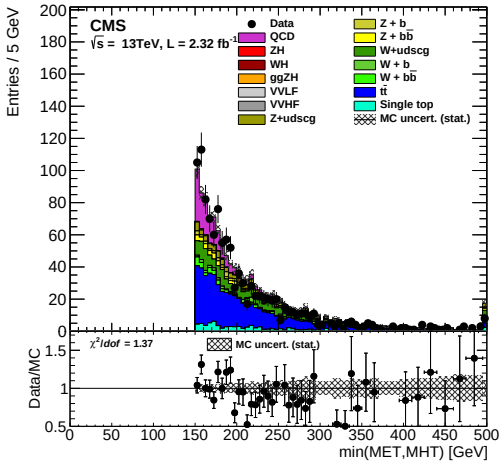
FIGURE C.9: Distributions of some variables for the Z+light-jets control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



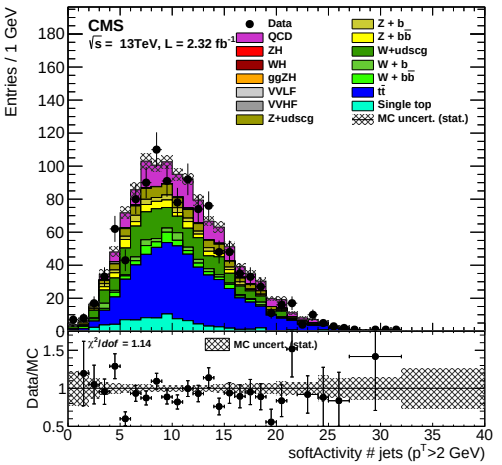
(A) m_{jj}



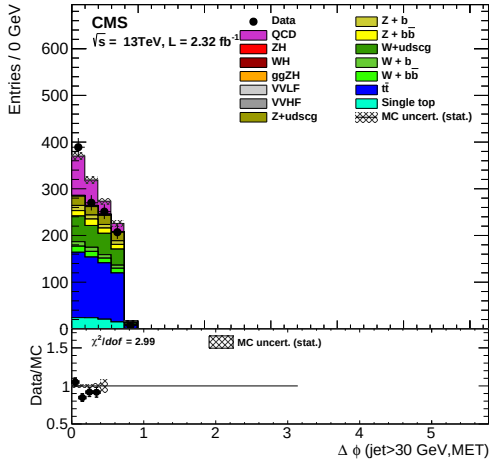
(B) min. Higgs jet p_T



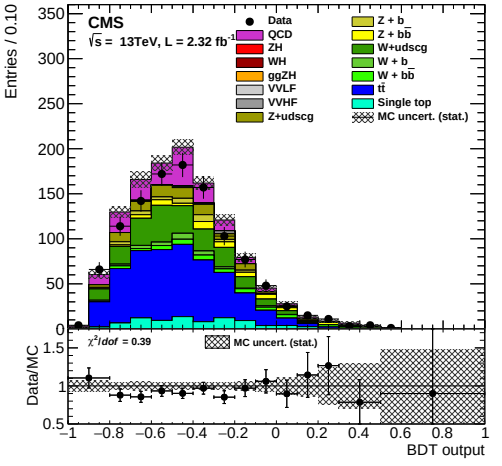
(C) MET



(D) Num. track jets $p_T > 2\,\text{GeV}$

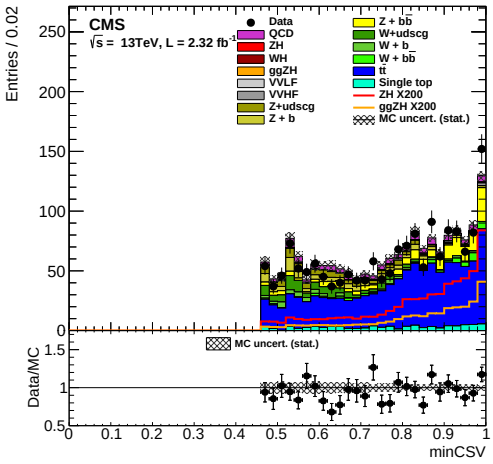


(E) $\min(\Delta\varphi(\text{jet}, \text{MET}))$

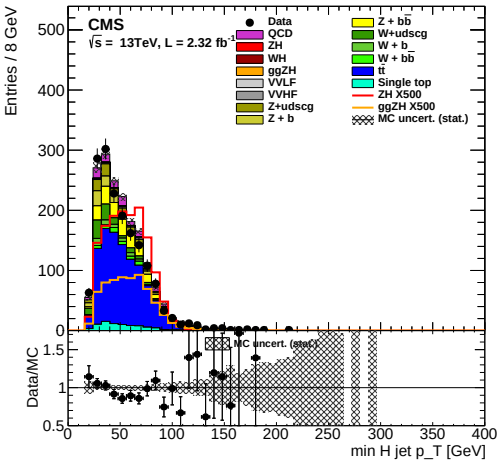


(F) BDT output

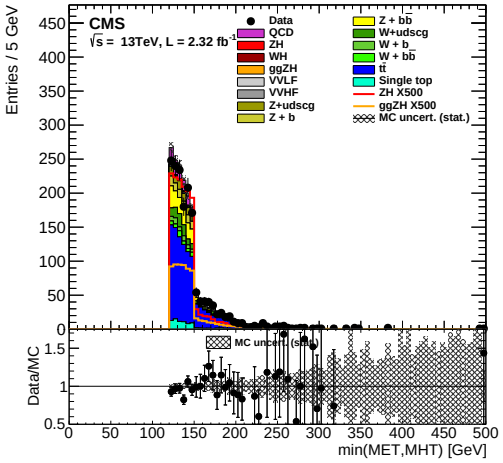
FIGURE C.10: Distributions of some variables for the multijet control region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



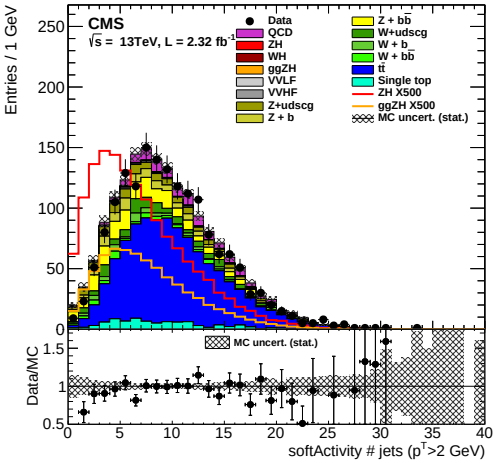
(A) $\mathrm{CSV}_{\mathrm{min}}$



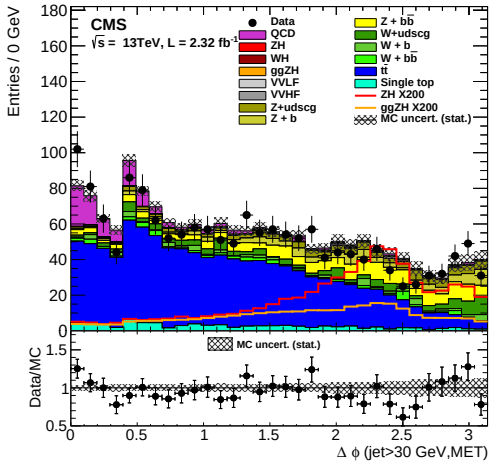
(B) min. Higgs jet p_T



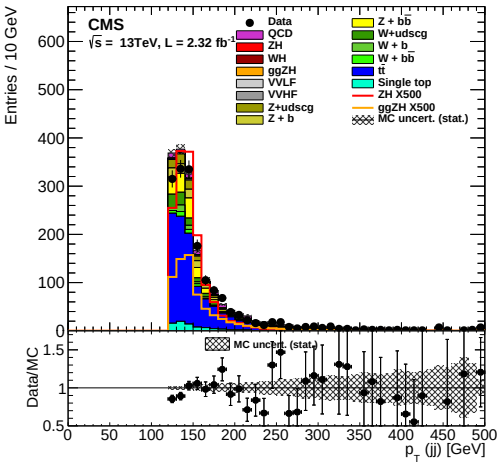
(C) MET



(D) Num. track jets $p_T > 2\,\mathrm{GeV}$

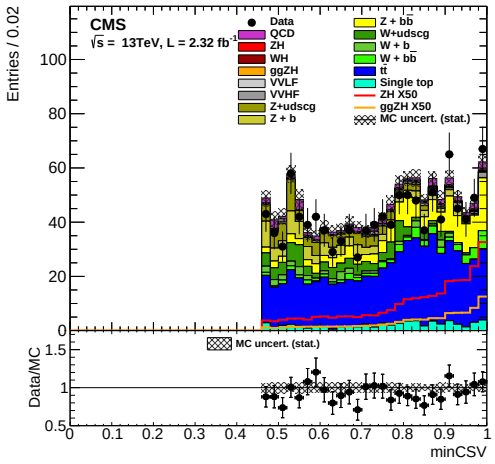


(E) $\min(\Delta\varphi(\mathrm{jet}, \mathrm{MET}))$

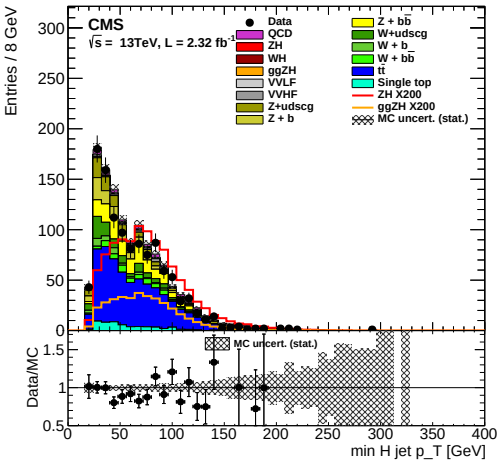


(F) Higgs candidate p_T

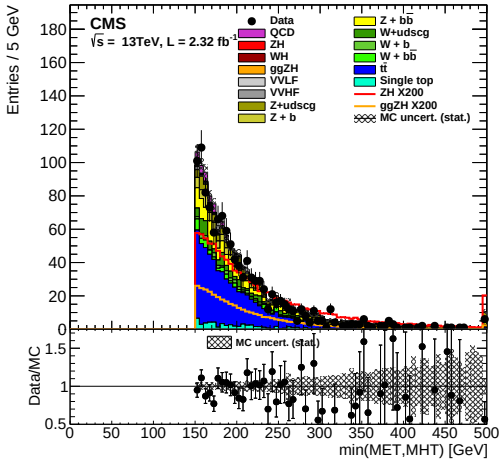
FIGURE C.11: Distributions of some variables for the signal region in the low $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.



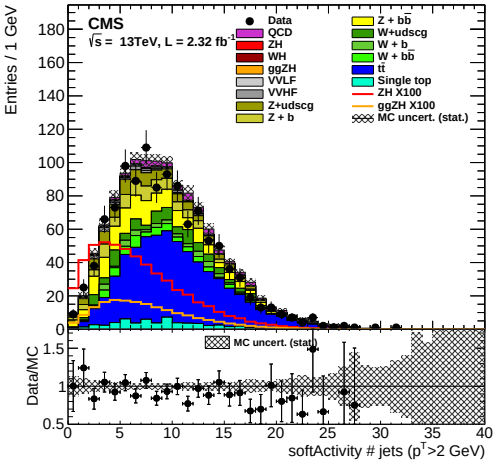
(A) $\mathrm{CSV}_{\mathrm{min}}$



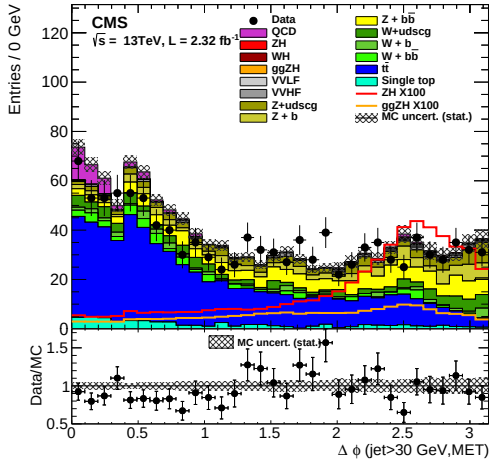
(B) min. Higgs jet p_T



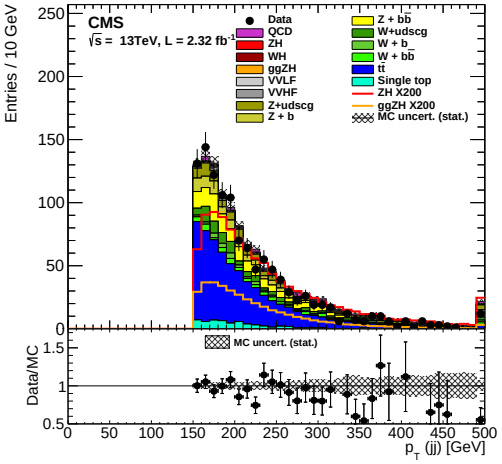
(C) MET



(D) Num. track jets $p_T > 2\,\mathrm{GeV}$



(E) $\min(\Delta\varphi(\mathrm{jet}, \mathrm{MET}))$



(F) Higgs candidate p_T

FIGURE C.12: Distributions of some variables for the signal region in the high $p_T(Z)$ bin analysis. The plots use the preliminary scale factors reported in Sect. 4.7.3.

Bibliography

[1] S. Glashow, "Partial Symmetries of Weak Interactions", *Nucl.Phys.* **22** (1961) 579–588.

[2] S. Weinberg, "A Model of Leptons", *Phys.Rev.Lett.* **19** (1967) 1264–1266.

[3] A. Salam in *Elementary Particle Physics*, N. Svartholm, ed., p. 367. Almquvist and Wiksell, Stockholm, 1968.

[4] Higgs, P. W. , "Broken symmetries, massless particles and gauge fields, *Phys. Lett.* **12**, 132–133 (1964).

[5] Englert, F. & Brout, R., "Broken symmetry and the mass of gauge vector mesons" *Phys. Lett.* **13**, 321–323 (1964).

[6] The CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3** S08004 (2008).

[7] The ATLAS Collaboration, "The ATLAS Experiment at the CERN Large Hadron Collider", *JINST* **3** (2008) S08003.

[8] Halzen and Martin, "Quarks and Leptons: An Introductory Course in Modern Particle Physics", *John Wiley & Sons* (1984).

[9] M. Peskin et al., "An Introduction to Quantum Field Theory", *Westview Press* (1995).

[10] C. Itzykson et J.B. Zuber, "Quantum Field Theory", *Dover Books on Physics* (2006).

[11] Troitsk Collaboration (V.N. Aseev et al.), "An upper limit on electron antineutrino mass from Troitsk experiment", *Phys.Rev.* **D84** (2011) 112003.

[12] Dirac, P. A. M. , "A Theory of Electrons and Protons", *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*,**126** (801): 360, (1930).

[13] C. N. Yang et R. L. Mills, "Isotopic Spin Conservation and a Generalized Gauge Invariance", *Phys. Rev.* **95**, 631(A) (1954).

[14] N. Cabibbo, "Unitary Symmetry and Leptonic Decays", *Phys. Rev. Lett.* **10**, 531 (1963).

[15] M. Kobayashi and T. Maskawa, "CP Violation in the Renormalizable Theory of Weak Interaction", *Prog. Theor. Phys.* **49**, 652 (1973).

[16] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 3. Higgs Properties".

[17] LHC Higgs Cross Section Working Group Collaboration, "Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables".

[18] Lyndon Evans and Philip Bryant (editors), "LHC Machine", *JINST* **3** S08001 (2008).

[19] CMS Collaboration, "Updated measurements of the Higgs boson at 125 GeV in the two photon decay channel", *CMS Physics Analysis Summary* **HIG-13-001**.

[20] ATLAS Collaboration, "Measurements of the properties of the Higgs-like boson in the two photon decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data", ATLAS-CONF-2013-012, ATLAS-COM-CONF-2013-015.

[21] CMS Collaboration, "Measurement of the properties of a Higgs boson in the four-lepton final state", *Phys. Rev. D* **89**, 092007 (2014).

[22] ATLAS Collaboration, "Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC", *Phys. Lett. B* **726**, 88 (2013).

[23] CMS Collaboration, "Update on the search for the standard model Higgs boson in pp collisions at the LHC decaying to W + W in the fully leptonic final state", *CMS Physics Analysis Summary* **HIG-13-003**.

[24] ATLAS Collaboration, "Measurements of the properties of the Higgs-like boson in the $WW^{(*)} \to \ell\nu\ell\nu$ decay channel with the ATLAS detector using 25 fb^{-1} of proton-proton collision data", ATLAS-CONF-2013-030, ATLAS-COM-CONF-2013-028.

[25] CMS Collaboration, "Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at sqrt(s) = 7 and 8 TeV", *CMS Physics Analysis Summary* **HIG-13-004**.

[26] [ATLAS collaboration], "Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector", ATLAS-CONF-**2013-108**, ATLAS-COM-CONF-**2013-095**.

[27] CMS Collaboration, "Search for the standard model Higgs boson produced in association with a W or a Z boson and decaying to bottom quarks", *Phys. Rev. D* **89**, 012003 (2014).

[28] ATLAS Collaboration, "Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector", ATLAS-**HIGG-2013-23**.

[29] CMS Collaboration, "Search for the standard model Higgs boson produced in vector boson fusion, and decaying to bottom quarks", *CMS Physics Analysis Summary* **HIG-13-011**.

[30] ATLAS Collaboration, "Search for the Standard Model Higgs boson produced in association with a vector boson and decaying to a b-quark pair with the ATLAS detector", *Phys. Lett. B* **718**, 369 (2012).

[31] CMS Collaboration, "Search for ttH events in the H $\rightarrow$ bb final state using the Matrix Element Method", *CMS Physics Analysis Summary* **HIG-14-010**.

[32] ATLAS collaboration, "Constraints on new phenomena via Higgs boson couplings and invisible decays with the ATLAS detector", *JHEP 11*, **206** (2015).

[33] CMS collaboration, "Search for invisible decays of a Higgs boson produced via vector boson fusion at $\sqrt{s} = 13\,\text{TeV}$. ", *CMS Physics Analysis Summary* **HIG-16-009**.

[34] ATLAS and CMS Collaborations, "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s} = 7$ and 8 TeV," ATLAS-CONF-**2015-044**.

[35] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults`

[36] The LHCb Collaboration, "The LHCb Detector at the LHC", *JINST* **3** (2008) S08005.

[37] The ALICE Collaboration, "The ALICE experiment at the CERN LHC", *JINST* **3** (2008) S08005.

[38] `http://indonesiaproud.files.wordpress.com/2010/02/lhc-experiments.jpg`

[39] CMS Collaboration, "The CMS experiment at the CERN LHC", *JINST* **3**, S08004 (2008).

[40] The CMS Collaboration, "Precise Mapping of the Magnetic Field in the CMS barrel Yoke using Cosmic Rays", *CMS Paper*, **CFT-09-015** (2010).

[41] CMS collaboration, "Description and performance of track and primary-vertex reconstruction with the CMS tracker", *JINST*, **9** P10009, (2014).

[42] `http://www.hephy.at/typo3temp/pics/911f43bdb2.png`

[43] The CMS Collaboration, "Performance and operation of the CMS electromagnetic calorimeter", *JINST*, **265** T03010 (2010).

[44] The CMS Collaboration, "Test beam results on the performance of the CMS electromagnetic calorimeter", *CMS Conference Report*, **2006/067**, (2006).

[45] V. D. Elvira, "Measurement of the Pion Energy Response and Resolution in the CMS HCAL Test Beam 2002 Experiment", *CMS Note*, **2004-020**, (2004).

[46] The CMS Collaboration, "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET", *CMS Physics Analysis Summary*, **PFT-09-001** (2009).

[47] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/PixelOfflineLHCC2015`

[48] R. Fruhwirth, "Application of Kalman filtering to track and vertex fitting", *Nucl. Instrum. Meth. A*,**262**, (1987).

[49] CMS collaboration, "Performance of CMS muon reconstruction in pp collision events at s = 7 TeV", 2012 *JINST* **7** P10002.

[50] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/TrackingPOGPlotsEPS2015`

[51] W. Waltenberger, R. Frühwirth, and P. Vanlaer, "Adaptive Vertex Fitting", *J. Phys. G* **34**, N343 (2007).

[52] M. Cacciari, G. Salam, and G. Soyez, "The anti-k t jet clustering algorithm", *JHEP 08*, **063** (2008).

[53] The CMS Collaboration, "Jet Performance in pp Collisions at s=7 Te, and MET", *CMS Physics Analysis Summary*, *CMS Physics Analysis Summary* **JME-10-003** (2010).

[54] The CMS Collaboration, "Missing transverse energy performance of the CMS detector", *JINST* **6** 09001 (2011).

[55] CMS Collaboration, "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV," *JINST* **10**, no. 06, P06005 (2015).

[56] CMS Collaboration, "Performance of Photon Reconstruction and Identification with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV," *JINST* **10**, no. 08, P08010 (2015).

[57] W. Adam, R. Fruhwirth, A. Strandlie et al., "Reconstruction of electrons with the Gaussian sum filter in the CMS tracker at LHC", *J. Phys, G*: Nucl. Part. Phys **31** (2005) N5–N20.

[58] CMS Collaboration, "The performance of the CMS muon detector in proton-proton collisions at sqrt(s) = 7 TeV at the LHC," *JINST* **8**, P11002 (2013).

[59] F. Beaudette , D. Benedetti , P. Janot , and M. Pioppi, "Electron Reconstruction within the Particle Flow Algorithm", *CMS Analysis Note*, **AN–2010–034**.

[60] A. Rizzi, F. Palla, G. Segneri, "Track impact parameter based b-tagging with CMS", *CMS Note*, CMS NOTE **2006/019**, (2006).

[61] The CMS Collaboration, "Identification of b quark jets at the CMS Experiment in the LHC Run 2", *CMS Physics Analysis Summary* **BTV-15-001**.

[62] The CMS Collaboration, "Performance of b tagging at s = 8 TeV in multijet, $t\bar{t}$ and boosted topology events", *CMS Physics Analysis Summary* **BTV-13-001**

[63] CMS Collaboration, "Identification of b-quark jets with the CMS experiment", *JINST* **8**, P04013 (2013).

[64] Y. Dokshitzer, G. Leder, S. Moretti, and B. Webber, "Better Jet Clustering Algorithms", *JHEP* **9708:001** (1997).

[65] CMS Collaboration, " Performance of b tagging in boosted topology events", *CMS Detector Performance Note*,**2014-031**, 2014.

[66] `http://www.pd.infn.it/%7Edorigo/l1trigger.jpg`

[67] `http://www.pd.infn.it/%7Edorigo/l1caltrigger.jpg`

[68] CMS Collaboration, "CMS Technical Design Report for the Level-1 Trigger Upgrade", CERN-LHCC-**2013-011**.

[69] The CMS Collaboration, "Performance of CMS Hadron Calorimeter Timing and Synchronization using Test Beam", *JINST* **5** (2010).

[70] CMS Collaboration, "Search for Higgs boson decaying to invisible particles and produced in association with a Z boson decaying to bottom quarks", *CMS Physics Analysis Summary* **HIG-13-028**.

[71] CMS Collaboration, "Search for narrow resonant pair production of the Higgs boson in the final state with four bottom quarks", *CMS Physics Analysis Summary* **HIG-14-013**.

[72] C. Vernieri, "Search for resonances decaying to H($b\bar{b}$) pairs with the CMS experiment at the LHC", Scuola Normale Superiore, Ph.D. Thesis (28 October 2014).

[73] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 13 TeV", *CMS Physics Analysis Summary* **HIG-16-002** (2016).

[74] CMS Collaboration, "Search for resonant pair production of Higgs bosons decaying to two bottom quark-antiquark pairs in proton-proton collisions at 13 TeV", *CMS Physics Analysis Summary* **HIG-16-002** (2016).

[75] B. Kniehl, *Phys. Rev.* **D42** (1990) 2253.

[76] S. Agostinelli et al. (GEANT4 Collaboration), *Nucl. Instrum. Methods Phys. Res.*, Sect. **A 506**, 250 (2003).

[77] S. Frixione, P. Nason, and C. Oleari, *J. High Energy Phys.* **11** (2007) 070.

[78] S. Frixione and B. R. Webber, *J. High Energy Phys.* **06** (2002) 029.

[79] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer, *J. High Energy Phys.* **06** (2011) 128.

[80] T. Sjotrand, S. Mrenna, and P. Z. Skands, **J. High Energy Phys. 05** (2006) 026.

[81] NNPDF Collaboration, R.D. Ball et al., Parton distributions for the LHC Run II. *JHEP* **1504**, 040 (2015).

[82] B. P. Roe, H.-J. Yang, J. Zhu, Y. Liu, I. Stancu, and G. McGregor, NIM: Sect.A **543**, 577 (2005).

[83] A. Hoecker, P. Speckmayer, J. Stelzer, J. Therhaag, E. von Toerne, and H. Voss, "TMVA: Toolkit for Multivariate Data Analysis," *PoS* A CAT **040** (2007).

[84] ATLAS and CMS Collaborations, "Procedure for the LHC higgs boson search combination in Summer 2011", ATL-PHYS-PUB-**2011-11**, CMS NOTE-**2011/005**, 2011.

[85] G. Petrucciani, "Observation of a new state in the search for the Higgs boson at CMS", Scuola Normale Superiore, Ph.D. Thesis (Academic year 2012/2013).

[86] A. L. Read, *J. Phys. G* **28**, 2693 (2002).

[87] T. Junk, *Nucl. Instrum. Methods Phys. Res.*, Sect. A **434**, 435 (1999).

[88] ATLAS and CMS Collaborations, LHC Higgs Combination Group, Reports No. ATL-PHYS-PUB 011-11 and No. CMS NOTE **2011-005**, 2011.

[89] J. B. MacQueen, "Some Methods for classification and Analysis of Multivariate Observations", *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, **1**:281-297 (1967).

[90] `https://apandre.files.wordpress.com/2011/08/kmeansclustering.jpg`

[91] `https://twiki.cern.ch/twiki/bin/view/CMSPublic/HighPtTrackingDP`